GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 8, 2003, 16:40:45 ; Search time 2182 Seconds
(without alignments)
11281.916 Million cell updates/sec

Title: US-09-847-665-4
Perfect score: 1520
Sequence: 1 acttaaaggaaaaatttatc.....cgtgtacggtccttgaaaa 1520

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	378.2	24.9	391	9	AI694972	AI694972 we44h07.x
2	345.4	22.7	367	9	AA324132	AA324132 EST27025
3	327	21.5	474	9	AL120587	AL120587 DKFZp761E
4	299.2	19.7	741	13	BI457956	BI457956 603198949
5	299.2	19.7	785	12	BG714610	BG714610 602676733
6	299.2	19.7	798	12	BG716320	BG716320 602676585
7	261.6	17.2	586	13	BI602678	BI602678 603252017
8	164.4	10.8	459	14	BQ601726	BQ601726 MI-P-HO-a
9	82.2	5.4	1101	17	CNS00EVL	AL069706 Drosophil
10	76.8	5.1	1101	17	CNS0039G	AL063921 Drosophil
11	75.4	5.0	905	17	CNS00KHX	AL077798 Drosophil
12	74.2	4.9	1101	17	CNS00EO7	AL069440 Drosophil
13	74	4.9	1101	17	CNS003B0	AL064078 Drosophil
c 14	73.6	4.8	975	17	BH179465	BH179465 014_P_10-
15	72.8	4.8	1101	17	CNS004ZW	AL055440 Drosophil
c 16	72.4	4.8	1167	17	CNS07360	AL427102 clone BA0
c 17	71.8	4.7	524	17	CNS01U9O	AL167541 Tetraodon
c 18	71.6	4.7	987	17	CNS014PQ	AL104456 Drosophil
c 19	71.4	4.7	1190	17	CNS02ON7	AL206908 Tetraodon
20	70.4	4.6	945	17	CNS04D0K	AL285149 Tetraodon
c 21	69.6	4.6	576	17	CNS035N7	AL228940 Tetraodon
22	69.2	4.6	1001	17	CNS0155H	AL105023 Drosophil
23	68.8	4.5	963	17	CNS00A4L	AL054918 Drosophil
c 24	68.4	4.5	1029	17	CNS01ZGM	AL174271 Tetraodon
c 25	68	4.5	787	17	AQ7404471	AQ7404471 HS_5504_A
26	67.8	4.5	887	12	BF274559	BF274559 GA__Eb002
27	67.8	4.5	1101	17	CNS00LT2	AL078714 Drosophil
28	67.6	4.4	876	17	CNS009G1	AL053529 Drosophil
c 29	67.6	4.4	1101	17	CNS00EO7	AL069440 Drosophil
30	67.4	4.4	483	9	AU087226	AU087226 AU087226
c 31	67.4	4.4	1101	17	CNS004ZW	AL055440 Drosophil
c 32	67.4	4.4	1225	17	CNS0161D	AL106171 Drosophil
c 33	67	4.4	942	17	CNS018GS	AL109318 Drosophil
34	67	4.4	970	17	CNS0182A	AL108796 Drosophil
35	67	4.4	1092	17	CNS020K7	AL175696 Tetraodon
36	67	4.4	1190	17	CNS02ON7	AL206908 Tetraodon
c 37	66.8	4.4	1027	17	CNS02T50	AL212733 Tetraodon
c 38	66.6	4.4	897	17	CNS07ABZ	AL436389 T7 end of
39	66.6	4.4	928	17	CNS00DKY	AL071865 Drosophil
c 40	66.6	4.4	928	17	CNS00DKY	AL071865 Drosophil
c 41	66.6	4.4	1101	17	CNS000B8	AL063632 Drosophil
c 42	66.6	4.4	1101	17	CNS00EVL	AL069706 Drosophil
c 43	66.6	4.4	1168	17	CNS06HDF	AL398857 T7 end of
44	66.4	4.4	996	17	CNS00FUH	AL071063 Drosophil
45	66.4	4.4	1135	17	CNS033GQ	AL226115 Tetraodon

ALIGNMENTS

RESULT 1
AI694972
LOCUS AI694972 391 bp mRNA linear EST 18-DEC-1999
DEFINITION we44h07.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2343997 3', mRNA sequence.
ACCESSION AI694972
VERSION AI694972.1 GI:4982872
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 391)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 672 Std Error: 0.00
Seq primer: -40UP from Gibco

```
          High quality sequence stop: 361.
FEATURES             Location/Qualifiers
     source          1. .391
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2343997"
                     /clone_lib="NCI_CGAP_Lu24"
                     /tissue_type="carcinoid"
                     /lab_host="DH10B"
                     /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                     modified polylinker; Plasmid DNA from the normalized
                     library NCI_CGAP_Lu5 was prepared, and ss circles were
                     made in vitro. Following HAP purification, this DNA was
                     used as tracer in a subtractive hybridization reaction.
                     The driver was PCR-amplified cDNAs from a pool of 5,000
                     clones made from the same library (cloneIDs
                     1414920-1417991 and 1520904-1522439). Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT       66 a    123 c     96 g    106 t
ORIGIN

  Query Match           24.9%;  Score 378.2;  DB 9;  Length 391;
  Best Local Similarity  99.2%;  Pred. No. 3.3e-75;
  Matches  380;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    1138 TATTTGCACCTAGAAATTTTAGGCAAGTGATAGCTGCGTAATCATACTGCGGCACCGTTT 1197
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TTTTTGCACCTAGAAATTTTAGGCAAGTGATAGCTGCGTAATCATACTGCGGCACCGTTT 60

Qy    1198 TTTTCTTGCAGCAGTAGCTGCTTGCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCT 1257
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TTTTCTTGCAGCAGTAGCTGCTTGCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCT 120

Qy    1258 CCGGCTCTCTCCTGCAGGATCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAG 1317
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CCGGCTCTCTCCTGCAGGATCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAG 180

Qy    1318 GTCGCCACTGCTTCAGTCCGGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAGCAGC 1377
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GTCGCCACTGCTTCAGTCCGGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAGCAGC 240

Qy    1378 CTGGATCGTCGTTCCCTCAGTCCGGACGCCACTGCTAGGTCCGACCACCGCCGCTTCTGA 1437
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CTGGATCGTCGTTCCCTCAGTCCGGACGCCACTGCTAGGTCCGACCACCGCCGCTTCTGA 300

Qy    1438 TATTTCGGTGAGTCTTTTCCTGTGGAGGTTTGGTCTCCCGATCTCTGTGGTAGCCACCTT 1497
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||| |||
Db     301 TATTTCGGTGAGTCTTTTCCTGTGGAGGTTTGGTCTCCCAATCTCTGTGGTAGCCAACTT 360

Qy    1498 AGGCGTGTACGGTCCTTTGAAAA 1520
           |||||||||||||||||||||||
Db     361 AGGCGTGTACGGTCCTTTGAAAA 383


RESULT 2
AA324132
LOCUS       AA324132                367 bp    mRNA    linear   EST 20-APR-1997
DEFINITION  EST27025 Cerebellum II Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA324132
VERSION     AA324132.1  GI:1976479
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 367)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
            ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
            , Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
            Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
            Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .367
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):124708"
                     /db_xref="taxon:9606"
                     /clone_lib="Cerebellum II"
                     /tissue_type="cerebellum"
                     /dev_stage="adult"
                     /note="Organ: brain; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"
BASE COUNT       54 a    113 c     92 g     98 t      10 others
ORIGIN

  Query Match           22.7%;  Score 345.4;  DB 9;  Length 367;
  Best Local Similarity  96.9%;  Pred. No. 8.9e-68;
  Matches  346;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy    1164 GTGATAGCTGCGTAATCATACTGCGGCACCGTTTTTTTTCTTGCAGCAGTAGCTGCTTTGCG 1223
           ||||||||||||||||||||||||||||||||||||||||| |||||||||||| |||||
Db       1 GTGATAGCTGCGTAATCATACTGCGGCACCGTTTTTTTTNTTGCAGCAGTAGCTGTTTGCG 60

Qy    1224 GAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCGGCTCTCTCCTGCAGGATCGGTCA 1283
           |||||||||||||||||||||||||||||||||||||| | ||||| |||||||||||||
Db      61 GAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCGGCTCTNTNCTGCAGGNTCGGTCA 120

Qy    1284 ACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGTCGCCACTGCTTCAGTCCGGTTCTC 1343
           |||||||||||||||||||||||||||||||||||||||||| |||| ||||||||||||
Db     121 ACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGTCGCCACTGNTTCANTCCGGTTCTC 180

Qy    1344 AAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTGGATCGTCGTTCCCTCAGTCCGGA 1403
           |||||||||||||||||||||||||||||||||||||| |||| |||||||||||||||
Db     181 AAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTGGNTCGTNGTTCCCTCAGTCCGGA 240

Qy    1404 CGCCACTGCTAGGTCCGACCACCGCCGCTTCTGATATTTCGGTGAGTCTTTTCCTGTGGA 1463
           |||||||||||||||||||||||||||||||| |||||| ||||||||||||||||||||
Db     241 CGCCACTGCTAGGTCCGACCACCGNCGCTTCTNATATTTCGGTGAGTCTTTTCCTGTGGA 300

Qy    1464 GGTTTGGTCTCCCGATCTCTGTGGTAGCCACCTTAGGCGTGTACGGTCCTTTGAAAA 1520
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GGTTTGGTCTCCCGATCTCTGTGGTAGCCACCTTAGGCGTGTACGGTCCTTTGAAAA 357


RESULT 3
AL120587
LOCUS       AL120587                474 bp    mRNA    linear   EST 25-FEB-2000
DEFINITION  DKFZp761E219_r1 761 (synonym: hamy2) Homo sapiens cDNA clone
            DKFZp761E219 5', mRNA sequence.
ACCESSION   AL120587
```

```
VERSION     AL120587.1  GI:5926486
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 474)
  AUTHORS   Blum,H., Bauersachs,S., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
  TITLE     EST (Blum, et al.)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Blum H
            MIPS
            Am Klopferspitz 18a D-82152 Martinsried, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by LMU (Ludwig Maximilians University,
            Munich/Germany) within the cDNA sequencing consortium of the German
            Genome Project.
            s1 sequence also available.
            This clone (DKFZp761E219) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .474
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp761E219"
                     /clone_lib="761 (synonym: hamy2)"
                     /tissue_type="amygdala"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
BASE COUNT       84 a    138 c    141 g    111 t
ORIGIN

  Query Match             21.5%;  Score 327;  DB 9;  Length 474;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-63;
  Matches  327;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1194 GTTTTTTTCTTGCAGCAGTAGCTGCTTGCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCA 1253
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     13 GTTTTTTTCTTGCAGCAGTAGCTGCTTGCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCA 72

Qy   1254 GTCTCCGGCTCTCTCCTGCAGGATCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAGC 1313
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     73 GTCTCCGGCTCTCTCCTGCAGGATCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAGC 132

Qy   1314 CCAGGTCGCCACTGCTTCAGTCCGGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAG 1373
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    133 CCAGGTCGCCACTGCTTCAGTCCGGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAG 192

Qy   1374 CAGCCTGGATCGTCGTTCCCTCAGTCCGGACGCCACTGCTAGGTCCGACCACCGCCGCTT 1433
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    193 CAGCCTGGATCGTCGTTCCCTCAGTCCGGACGCCACTGCTAGGTCCGACCACCGCCGCTT 252

Qy   1434 CTGATATTTCGGTGAGTCTTTTCCTGTGGAGGTTTGGTCTCCCGATCTCTGTGGTAGCCA 1493
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    253 CTGATATTTCGGTGAGTCTTTTCCTGTGGAGGTTTGGTCTCCCGATCTCTGTGGTAGCCA 312

Qy   1494 CCTTAGGCGTGTACGGTCCTTTGAAAA 1520
          |||||||||||||||||||||||||||
Db    313 CCTTAGGCGTGTACGGTCCTTTGAAAA 339


RESULT 4
BI457956
LOCUS       BI457956                  741 bp    mRNA    linear   EST 21-AUG-2001
DEFINITION  603198949F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5278356 5',
            mRNA sequence.
ACCESSION   BI457956
VERSION     BI457956.1  GI:15248612
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 741)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11703  row: a  column: 13
            High quality sequence stop: 741.
FEATURES             Location/Qualifiers
     source          1. .741
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5278356"
                     /clone_lib="NIH_MGC_96"
                     /tissue_type="hypothalamus"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.3 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length  clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      177 a    180 c    218 g    166 t
ORIGIN

  Query Match             19.7%;  Score 299.2;  DB 13;  Length 741;
  Best Local Similarity   99.0%;  Pred. No. 2.6e-57;
  Matches  301;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy   1217 GCTTGCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCGGCTCTCTCCTGCAGGA 1276
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6 GCTTGCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCGGCTCTCTCCTGCAGGA 65

Qy   1277 TCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGTCGCCACTGCTTCAGTCC 1336
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     66 TCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGTCGCCACTGCTTCAGTCC 125

Qy   1337 GGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTGGATCGTCGTTCCCTCA 1396
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    126 GGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTGGATCGTCGTTCCCTCA 185

Qy   1397 GTCCGGACGCCACTGCTAGGTCCGACCACCGCCGCTTCTGATATTTCGGTGAGTCTTTTC 1456
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    186 GTCCGGACGCCACTACTAGGTCCGACCACCGCCGCTTCTGATATTTCGGTGAGTCTTTTC 245

Qy   1457 CTGTGGAGGTTTGGTCTCCCGATCTCTGTGGTAGCCACCTTAGGCGTGTACGGTCCTTTG 1516
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    246 CTGTGGAGGTTTGGTCTCCCGATCCCTGTGGTAGCCACCTTAGGCGTGTACGGTCCTTTG 305

Qy   1517 AAAA 1520
          ||||
Db    306 AAAA 309


RESULT 5
BG714610
LOCUS       BG714610                  785 bp    mRNA    linear   EST 08-MAY-2001
DEFINITION  602676733F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4799121 5',
```

mRNA sequence.
ACCESSION BG714610
VERSION BG714610.1 GI:13993541
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 785)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM10688 row: a column: 10
High quality sequence stop: 777.

FEATURES Location/Qualifiers
source 1. .785
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4799121"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.3 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 196 a 188 c 224 g 177 t
ORIGIN

Query Match 19.7%; Score 299.2; DB 12; Length 785;
Best Local Similarity 99.0%; Pred. No. 2.6e-57;
Matches 301; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy   1217 GCTTGCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCGGCTCTCTCCTGCAGGA 1276
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6 GCTTGCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCGGCTCTCTCCTGCAGGA 65

Qy   1277 TCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGTCGCCACTGCTTCAGTCC 1336
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     66 TCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGTCGCCACTGCTTCAGTCC 125

Qy   1337 GGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTGGATCGTCGTTCCCTCA 1396
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    126 GGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTGGATCGTCGTTCACTCA 185

Qy   1397 GTCCGGACGCCACTGCTAGGTCCGACCACCGCCGCTTCTGATATTTCGGTGAGTCTTTTC 1456
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    186 GTCCGGACGCCACTACTAGGTCCGACCACCGCCGCTTCTGATATTTCGGTGAGTCTTTTC 245

Qy   1457 CTGTGGAGGTTTGGTCTCCCGATCTCTGTGGTAGCCACCTTAGGCGTGTACGGTCCTTTG 1516
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    246 CTGTGGAGGTTTGGTCTCCCGATCCCTGTGGTAGCCACCTTAGGCGTGTACGGTCCTTTG 305

Qy   1517 AAAA 1520
          ||||
Db    306 AAAA 309
```

RESULT 6
BG716320
LOCUS BG716320 798 bp mRNA linear EST 08-MAY-2001
DEFINITION 602676585F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4798964 5', mRNA sequence.
ACCESSION BG716320
VERSION BG716320.1 GI:13995507
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 798)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM10687 row: j column: 21
High quality sequence stop: 713.

FEATURES Location/Qualifiers
source 1. .798
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4798964"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.3 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 195 a 195 c 229 g 179 t
ORIGIN

Query Match 19.7%; Score 299.2; DB 12; Length 798;
Best Local Similarity 99.0%; Pred. No. 2.6e-57;
Matches 301; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy   1217 GCTTGCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCGGCTCTCTCCTGCAGGA 1276
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      6 GCTTGCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCGGCTCTCTCCTGCAGGA 65

Qy   1277 TCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGTCGCCACTGCTTCAGTCC 1336
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     66 TCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGTCGCCACTGCTTCAGTCC 125

Qy   1337 GGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTGGATCGTCGTTCCCTCA 1396
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    126 GGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTGGATCGTCGTTCACTCA 185

Qy   1397 GTCCGGACGCCACTGCTAGGTCCGACCACCGCCGCTTCTGATATTTCGGTGAGTCTTTTC 1456
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    186 GTCCGGACGCCACTACTAGGTCCGACCACCGCCGCTTCTGATATTTCGGTGAGTCTTTTC 245

Qy   1457 CTGTGGAGGTTTGGTCTCCCGATCTCTGTGGTAGCCACCTTAGGCGTGTACGGTCCTTTG 1516
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    246 CTGTGGAGGTTTGGTCTCCCGATCCCTGTGGTAGCCACCTTAGGCGTGTACGGTCCTTTG 305

Qy   1517 AAAA 1520
```

```
Db     306 AAAA 309


RESULT 7
BI602678
LOCUS       BI602678                586 bp    mRNA    linear   EST 07-SEP-2001
DEFINITION  603252017F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5303692 5',
            mRNA sequence.
ACCESSION   BI602678
VERSION     BI602678.1  GI:15495617
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 586)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
  COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11769  row: a  column: 05
            High quality sequence stop: 586.
FEATURES             Location/Qualifiers
     source          1. .586
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5303692"
                     /clone_lib="NIH_MGC_96"
                     /tissue_type="hypothalamus"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.3 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length  clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      129 a    148 c    182 g    127 t
ORIGIN

  Query Match           17.2%;  Score 261.6;  DB 13;  Length 586;
  Best Local Similarity   98.5%;  Pred. No. 8.6e-49;
  Matches  264;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy   1253 AGTCTCCGGCTCTCTCCTGCAGGATCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAG 1312
           ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AGCGGCCGGCTCTCTCCTGCAGGATCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAG 60

Qy   1313 CCCAGGTCGCCACTGCTTCAGTCCGGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGA 1372
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CACAGGTCGCCACTGCTTCAGTCCGGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGA 120

Qy   1373 GCAGCCTGGATCGTCGTTCCCTCAGTCCGGACGCCACTGCTAGGTCCGACCACCGCCGCT 1432
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCAGCCTGGATCGTCGTTCCCTCAGTCCGGACGCCACTGCTAGGTCCGACCACCGCCGCT 180

Qy   1433 TCTGATATTTCGGTGAGTCTTTTCCTGTGGAGGTTTGGTCTCCCGATCTCTGTGGTAGCC 1492
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TCTGATATTTCGGTGAGTCTTTTCCTGTGGAGGTTTGGTCTCCCGATCTCTGTGGTAGCC 240

Qy   1493 ACCTTAGGCGTGTACGGTCCTTTGAAAA 1520
```

```
Db     241 ACCTTAGGCGTGTACGGTCCTTTGAAAA 268


RESULT 8
BQ601726
LOCUS       BQ601726                459 bp    mRNA    linear   EST 24-JUN-2002
DEFINITION  MI-P-HO-afw-d-02-1-UM.s1 MI-P-HO Sus scrofa cDNA clone
            MI-P-HO-afw-d-02-1-UM 3', mRNA sequence.
ACCESSION   BQ601726
VERSION     BQ601726.1  GI:21548452
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 459)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
  COMMENT   Contact: Tuggle CK
            Molecular Genetics Laboratory, Department of Animal Science
            Iowa State University
            201 Kildee Hall, Ames, IA 50011-3150, USA
            Tel: 5152944252
            Fax: 5152942401
            Email: cktuggle@iastate.edu
            Tissue Procurement: Dr. Chris Tuggle, Iowa State University
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            Seq primer: M13 FORWARD
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .459
                     /organism="Sus scrofa"
                     /strain="crossbreed"
                     /db_xref="taxon:9823"
                     /clone="MI-P-HO-afw-d-02-1-UM"
                     /clone_lib="MI-P-HO"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-HO
                     library is a normalized library derived from hypothalamus
                     at estrus days 0 and 12 and ovary at estrus days 0, 5 and
                     12. For a detailed description of the library from which
                     this clone was derived, please visit our web site at
                     http://pigest.genome.iastate.edu/.
                     TAG_LIB=MI-P-HO
                     TAG_TISSUE=d_0_hypothalamus
                     TAG_SEQ=TAGATG"
BASE COUNT       79 a    152 c    126 g    102 t
ORIGIN

  Query Match           10.8%;  Score 164.4;  DB 14;  Length 459;
  Best Local Similarity   74.0%;  Pred. No. 8.8e-27;
  Matches  256;  Conservative    0;  Mismatches   71;  Indels   19;  Gaps    3;

Qy   1194 GTTTTTTTCTTGCAGCAGTAGCTGCTTGC-GGAGGAGGTCTGCCCACTGCAGCTCTCTGC 1252
           |||||||  |||||||||||  |||||||| ||||| | ||||||||||||| |||||||||
Db      1 GTTTTTTCCTTGCAGCAGCTGCTGCTTGCTGGAGGCGATCTGCCCACTGCGGCTCTCTGC 60

Qy   1253 AGTCTCCGGCTCTCTCCTGCAGGATCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAG 1312
           ||||||   ||||  |||||||||| |||||| | |||||||||||| ||||||||||
Db     61 AGTCTCTAGCTCGGTCCTGCAGGACCGGTCACGGAAGCCGTCGCCGCTCTCTGCACCCCA 120

Qy   1313 CCCAGGTCGCCACTGCTTCAGTCCGGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGA 1372
           || | | || |||| || || | ||| |   | |||||||||   |||   ||   |
Db    121 GCCCAGCCACCGCTGCCTCTGTGCCGTTTTTGTATCCTCAGCACAGCCTTCGTCCCTGCA 180
```

```
Qy   1373 GCAGCCTGGATCGTC-------GTTCCCTCAGTCCGGACGCCACTGCTAGGT-------- 1417
Db    181 GCAGCCTCGACTGTCGCTCCAGGCTCCCACAACCCGACCGCTACCGCTCGTTCCACCTCT 240

Qy   1418 ---CCGACCACCGCCGCTTCTGATATTTCGGTGAGTCTTTTCCTGTGGAGGTTTGGTCTC 1474
Db    241 GCCCCAGCCACCGCCGCTTCTGATATTCCGGTGAGTCTTCCCTTGCGGAGGTCAGGTCTC 300

Qy   1475 CCGATCTCTGTGGTAGCCACCTTAGGCGTGTACGGTCCTTTGAAAA 1520
Db    301 CCGATCTTGGCGGTAGCCATCTTAGGCTTGTGAGGTCCTTTGAAAA 346
```

```
RESULT 9
CNS00EVL
LOCUS       CNS00EVL               1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP):
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1..1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR29B23"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
BASE COUNT      419 a     91 c     60 g    299 t    232 others
ORIGIN

  Query Match             5.4%;  Score 82.2;  DB 17;  Length 1101;
  Best Local Similarity   35.8%;  Pred. No. 3.7e-08;
  Matches  150;  Conservative   78;  Mismatches  191;  Indels    0;  Gaps    0;

Qy      1 ACTTAAAGGAAAAATTTATCTATAAACTGACAGAATTTAGAAATAAATACAACAATATGT 60
Db    680 ATTTWAAATAWAAAAAAAAAAAAAAAWATWAAWAATWATAWATAAWTTAAAAWAATAAAA 739

Qy     61 AAACAGTTTTAATATCTGTGATAGTAACAAATTCTTTAAATCTGGAAAATAATAGTCACT 120
Db    740 WAAWAATWAWWATAATATWWATATATAWTTWTAWWWATWWAWWWTATAWATAWAATAWAA 799

Qy    121 TAAAATTTTAAAAAATTGTTCAATTAATAAATGATCCAAGTTAGAAATATGAACAAAATA 180
Db    800 WAWAWATAAATAWATAWATWAAAWAWAWATAWWATWATATAWWAATAWAWAAAAAATWTA 859

Qy    181 AACCTCACCAATAATTACTATAGAGAGGAAATTTTAATTACTGCAAAGCTTTCCATCCTA 240
Db    860 ATATWAATWATAAWAAAAAATAWAWTTWWTWTTTTWWAWWATATAAAWAWATAWAAAAWAA 919

Qy    241 TAAATACATTATCAAATAGTTTAACCATTTCTTTAATGCTGAGATTTAGATTATTTCCAA 300
Db    920 AAAAAAAAAAATAAWAWWWTWATATTWTTATTAAAWTWTATWWATTWATTWWAWTWTATA 979

Qy    301 TTAACTCAAAAGCATCAAGCAAATGTTATGATTTCTAAGAATAAACATAACTTTCCATTT 360
Db    980 WTTWTATWTATATATWWTAWTAWATATATTTATTAAWWTATATTTTAWAAAWTAATATAT 1039

Qy    361 TGGCTTTTGTATATATGTATATTTCTAACGGCTGTTAAAGCCAGCATTAAGAAGGAGAA 419
Db   1040 WATAWWWTAWATATAWAWWAATTAWTTATATATWTAAWATAWWAAAAAWAAAWAWATAW 1098
```

```
RESULT 10
CNS0039G
LOCUS       CNS0039G               1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL063921
VERSION     AL063921.1  GI:4941778
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1..1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR08K10"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      201 a     64 c    131 g    202 t    503 others
ORIGIN

  Query Match             5.1%;  Score 76.8;  DB 17;  Length 1101;
  Best Local Similarity   17.2%;  Pred. No. 6.2e-07;
  Matches  116;  Conservative  291;  Mismatches  266;  Indels    3;  Gaps    1;

Qy      3 TTAAAGGAAAAATTTATCTATAAACTGACAGAATTTAGAAATAAATACAACAATATGTAA 62
Db    413 TTTAWAAAWAAAATAATTWWAAWAWAAAAAATTWWAAAAWAAAAWAWTAWWTTTAWTWAA 472

Qy     63 ACAGTTTTAATATCTGTGATAGTAACAAATTCTTTAAATCTGGAAAATAATAGTCACTTA 122
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 8, 2003, 16:37:35 ; Search time 351 Seconds
(without alignments)
9752.243 Million cell updates/sec

Title: US-09-847-665-4
Perfect score: 1520
Sequence: 1 acttaaaggaaaaatttatc.....cgtgtacggtcctttgaaaa 1520

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

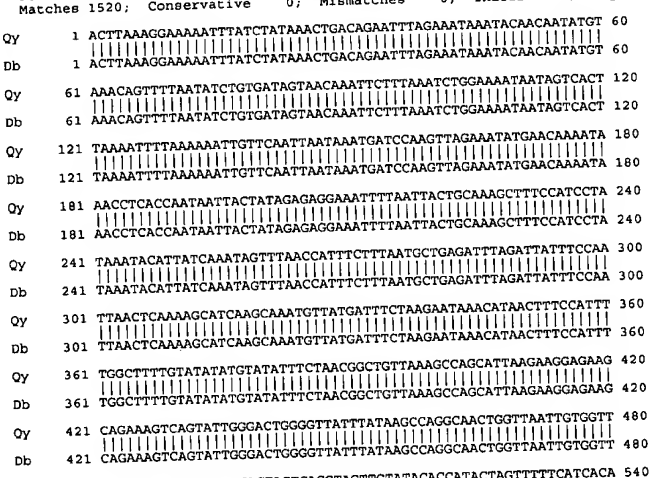
Result No.	Score	% Query Match	Length	DB	ID	Description
1	1520	100.0	1520	24	ABK10803	Human nucleosomal
2	1520	100.0	3699	24	ABK10805	DNA encoding human
3	522.4	34.4	2720	24	ABK10802	Human nucleosomal
4	380.6	25.0	629	21	AAC01166	Human secreted pro
5	233.8	15.4	1725	24	ABK10360	Murine nucleosomal
6	166.8	11.0	2819	24	ABK10801	Murine nucleosomal
7	67	4.4	513445	22	AAI61373	Soybean 318013 reg
c 8	65.2	4.3	17967	24	ABL33015	Human immune syste
c 9	64.8	4.3	6419	24	ABL32267	Human immune syste
10	64.4	4.2	19124	18	AAT72882	Plasmodium var-7 g
11	64.4	4.2	19124	21	AAZ98287	Plasmodium var-7 p
c 12	63.8	4.2	83391	24	ABQ67093	Human angiogenesis
13	63	4.1	3991	22	AAD16633	Human novel protei
c 14	63	4.1	6286	22	AAS46591	Tumour suppressor
c 15	63	4.1	17234	24	ABQ67018	Human angiogenesis
c 16	63	4.1	18817	24	ABL70162	Chemically treated
c 17	63	4.1	18817	24	ABL34495	Human metastasis a
c 18	62.8	4.1	7849	24	ABL92278	Chemically treated
c 19	62.8	4.1	7849	24	AAD22329	Chemically treated
c 20	62.6	4.1	7597	24	ABL33013	Human immune syste
c 21	62.6	4.1	16217	24	ABL32624	Human immune syste
c 22	62.4	4.1	6109	24	ABL32326	Human immune syste
c 23	62.4	4.1	6109	24	AAS61077	Human gene regulat
c 24	62.4	4.1	7814	22	AAS46530	Tumour suppressor
c 25	62.2	4.1	6317	24	ABL49311	Human polynucleoti
c 26	62.2	4.1	6317	24	ABL32408	Human immune syste
c 27	62	4.1	5875	24	ABL32289	Human immune syste
c 28	61.8	4.1	17594	24	ABL34026	Human immune syste
c 29	61.4	4.0	11092	24	ABL33513	Human immune syste
c 30	61	4.0	7346	24	ABL32345	Human immune syste
c 31	60.8	4.0	6175	24	ABL33307	Human immune syste
c 32	60.8	4.0	37515	24	ABQ66998	Human angiogenesis
c 33	60.6	4.0	6082	24	ABL33753	Human immune syste
c 34	60.6	4.0	7862	24	ABL34055	Human immune syste
c 35	60.6	4.0	8085	22	AAS46480	Tumour suppressor
c 36	60.6	4.0	8085	24	ABK33987	Human DNA for stag
c 37	60.6	4.0	16173	24	ABL34469	Human metastasis a
c 38	60.4	4.0	5267	24	ABQ67043	Human angiogenesis
c 39	60.4	4.0	6048	24	ABQ67002	Human angiogenesis
c 40	60.4	4.0	6306	22	AAS45516	Chemically pretrea
c 41	60.4	4.0	6306	24	ABK28458	DNA transcription
c 42	60.4	4.0	6370	24	ABL70567	Chemically treated
c 43	60.4	4.0	6370	24	ABK31348	Signal transductio
44	60.2	4.0	4590	7	AAN60472	Sequence encoding
c 45	60.2	4.0	6027	24	ABN80172	Human chemically m

ALIGNMENTS

RESULT 1
ABK10803
ID ABK10803 standard; DNA; 1520 BP.
XX
AC ABK10803;
XX
DT 21-MAY-2002 (first entry)
XX
DE Human nucleosomal assembly protein IL2 (NAP1L2 or BPX) promoter.
XX
KW Human nucleosomal assembly protein IL2; NAP1L2; BPX; nootropic;
KW neurotropic; cytostatic; neural system defect; cancer;
KW neuronal cell tumour; gene therapy; neurodegenerative disease;
KW Parkinson's disease; Alzheimer's disease; spina bifida; anencephaly;
KW brain structure; brain organisation; X-chromosome linked disorder;
KW nucleosome activity; cell cycle; cyclin; transcription factor control;
KW histone; promoter; ds.
XX
OS Homo sapiens.
XX
PN WO200185995-A2.
XX
PD 15-NOV-2001.
XX
PF 04-MAY-2001; 2001WO-IB00960.
XX
PR 05-MAY-2000; 2000US-202111P.
XX
PA (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Avner P, Rogner UC, Spyropoulos D, Rougeulle C;

XX
DR WPI; 2002-075254/10.
XX
PT Screening neural system defects in mammal, comprises detecting a
PT modification of NAP1L2 and/or Nap1l2 gene chromosomal material, that
PT causes a loss of biological function and correlating the modification
PT with the defect -
XX
PS Claim 24; Fig 10; 83pp; English.
XX
CC The invention describes a method of screening mammalian neural system
CC defects, comprising detecting a modification of the nucleosomal assembly
CC protein NAP1L2 and/or Nap1l2 gene in a human and/or mouse, where
CC modification is substitution, deletion, frame-shift, insertion aberrant
CC or altered epigenetic control, that causes a loss of function in the gene
CC and correlating the gene modification with a potential for a neural
CC system defect. The invention includes a neuronal cell containing a
CC recombinant chromosome on which Nap1L2 is located. The cell is useful for
CC screening therapeutic compounds by introducing a compound to be screened
CC and correlating change in the proliferation of cells with the activity of
CC the compound, where change in proliferating cells is preferably a control
CC of cancer of neural cells. The Nap1L2 promoter is useful for predicting
CC tumour of neural origin, and for genetic therapy of neurodegenerative
CC diseases such as Parkinson's, Alzheimer's diseases or accidents. Nap1L2
CC is also implicated in spina bifida with or without anencephaly, loss of
CC brain structure/organisation, X-chromosome linked disorders and
CC inappropriate control of nucleosome activity in neurons, neuronal cell
CC cycle, cyclins, DNA binding transcription factors, histones and histone
CC shuttling. This sequence represents the promoter of the human
CC nucleosomal assembly protein (NAP1L2 or BPX) gene, a homologue of the
CC murine nucleosomal assembly protein (Nap1L2 or Bpx) gene described in
CC the invention.
XX
SQ Sequence 1520 BP; 444 A; 321 C; 302 G; 453 T; 0 other;

Query Match 100.0%; Score 1520; DB 24; Length 1520;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1520; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



RESULT 2
ABK10805

```
ID   ABK10805 standard; DNA; 3699 BP.
XX
AC   ABK10805;
XX
DT   21-MAY-2002  (first entry)
XX
DE   DNA encoding human nucleosomal assembly protein (NAP1L2/BPX) protein.
XX
KW   Human nucleosomal assembly protein; NAP1L2; BPX; nootropic;
KW   neurotropic; cytostatic; neural system defect; cancer;
KW   neuronal cell tumour; gene therapy; neurodegenerative disease;
KW   Parkinson's disease; Alzheimer's disease; spina bifida; anencephaly;
KW   brain structure; brain organisation; X-chromosome linked disorder;
KW   nucleosome activity; cell cycle; cyclin; transcription factor control;
KW   histone; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1521..2906
FT                   /*tag=  a
FT                   /product= "NAP1L2"
FT                   /note= "Nucleosomal assembly protein IL2"
XX
PN   WO200185995-A2.
XX
PD   15-NOV-2001.
XX
PF   04-MAY-2001; 2001WO-IB00960.
XX
PR   05-MAY-2000; 2000US-202111P.
XX
PA   (INSP ) INST PASTEUR.
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Avner P,  Rogner UC,  Spyropoulos D,  Rougeulle C;
XX
DR   WPI; 2002-075254/10.
XX
PT   Screening neural system defects in mammal, comprises detecting a
PT   modification of NAP1L2 and/or Nap112 gene chromosomal material, that
PT   causes a loss of biological function and correlating the modification
PT   with the defect
XX
PS   Disclosure; Fig 13; 83pp; English.
XX
CC   The invention describes a method of screening mammalian neural system
CC   defects, comprising detecting a modification of the nucleosomal assembly
CC   protein NAP1L2 and/or Nap112 gene in a human and/or mouse, where
CC   modification is substitution, deletion, frame-shift, insertion aberrant
CC   or altered epigenetic control, that causes a loss of function in the gene
CC   and correlating the gene modification with a potential for a neural
CC   system defect. The invention includes a neuronal cell containing a
CC   recombinant chromosome on which Nap1L2 is located. The cell is useful for
CC   screening therapeutic compounds by introducing a compound to be screened
CC   and correlating change in the proliferation of cells with the activity of
CC   the compound, where change in proliferating cells is preferably a control
CC   of cancer of neural cells. The Nap1L2 promoter is useful for predicting
CC   tumour of neural origin, and for genetic therapy of neurodegenerative
CC   diseases such as Parkinson's, Alzheimer's diseases or accidents. Nap1L2
CC   is also implicated in spina bifida with or without anencephaly, loss of
CC   brain structure/organisation, X-chromosome linked disorders and
CC   inappropriate control of nucleosome activity in neurons, neuronal cell
CC   cycle, cyclins, DNA binding transcription factors, histones and histone
CC   shuttling. This sequence represents the human nucleosomal assembly
CC   protein (NAP1L2 or BPX) gene, a homologue of the murine nucleosomal
CC   assembly protein (Nap1L2 or Bpx) gene described in the invention.
XX
SQ   Sequence 3699 BP; 1144 A; 636 C; 805 G; 1114 T; 0 other;

Query Match             100.0%;  Score 1520;  DB 24;  Length 3699;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 1520;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy       1 ACTTAAAGGAAAAATTTATCTATAAACTGACAGAATTTAGAAATAAATACAACAATATGT 60
Db       1 ACTTAAAGGAAAAATTTATCTATAAACTGACAGAATTTAGAAATAAATACAACAATATGT 60

Qy      61 AAACAGTTTTAATATCTGTGATAGTAACAAATTCTTTAAATCTGGAAAATAATAGTCACT 120
Db      61 AAACAGTTTTAATATCTGTGATAGTAACAAATTCTTTAAATCTGGAAAATAATAGTCACT 120

Qy     121 TAAAATTTTAAAAAATTGTTCAATTAATAAATGATCCAAGTTAGAAATATGAACAAAATA 180
Db     121 TAAAATTTTAAAAAATTGTTCAATTAATAAATGATCCAAGTTAGAAATATGAACAAAATA 180

Qy     181 AACCTCACCAATAATTACTATAGAGAGGAAATTTTAATTACTGCAAAGCTTTCCATCCTA 240
Db     181 AACCTCACCAATAATTACTATAGAGAGGAAATTTTAATTACTGCAAAGCTTTCCATCCTA 240

Qy     241 TAAATACATTATCAAATAGTTTAACCATTTCTTTAATGCTGAGATTTAGATTATTTCCAA 300
Db     241 TAAATACATTATCAAATAGTTTAACCATTTCTTTAATGCTGAGATTTAGATTATTTCCAA 300

Qy     301 TTAACTCAAAAGCATCAAGCAAATGTTATGATTTCTAAGAATAAACATAACTTTCCATTT 360
Db     301 TTAACTCAAAAGCATCAAGCAAATGTTATGATTTCTAAGAATAAACATAACTTTCCATTT 360

Qy     361 TGGCTTTTGTATATATGTATATTTCTAACGGCTGTTAAAGCCAGCATTAAGAAGGAGAAG 420
Db     361 TGGCTTTTGTATATATGTATATTTCTAACGGCTGTTAAAGCCAGCATTAAGAAGGAGAAG 420

Qy     421 CAGAAAGTCAGTATTGGGACTGGGGTTATTTATAAGCCAGGCAACTGGTTAATTGTGGTT 480
Db     421 CAGAAAGTCAGTATTGGGACTGGGGTTATTTATAAGCCAGGCAACTGGTTAATTGTGGTT 480

Qy     481 AATTGTCTGGTATGTTTACTAGTCACGTAGTTGTATACACCATACTAGTTTTTCATCACA 540
Db     481 AATTGTCTGGTATGTTTACTAGTCACGTAGTTGTATACACCATACTAGTTTTTCATCACA 540

Qy     541 GGCCCTCATTCGCCCCCACTGCCATCGGACTTCCTCCTCCTCCCCTCACAGGAAATGTTT 600
Db     541 GGCCCTCATTCGCCCCCACTGCCATCGGACTTCCTCCTCCTCCCCTCACAGGAAATGTTT 600

Qy     601 CGAGAATTTTTCAACCTAAAATCATATAGCTTGTGAAAAATACCGACAAACATAATATAG 660
Db     601 CGAGAATTTTTCAACCTAAAATCATATAGCTTGTGAAAAATACCGACAAACATAATATAG 660

Qy     661 AATATTTAAATAACTGACACGCCACCTAAAGACCATCAGTGCTAATTCCTGGTGTTTTTA 720
Db     661 AATATTTAAATAACTGACACGCCACCTAAAGACCATCAGTGCTAATTCCTGGTGTTTTTA 720

Qy     721 ATCTTTGAAGCGTTTGTTTATCAGCTCTTCCACCATCCACCTCTCCCCTCCCCAGGTCCC 780
Db     721 ATCTTTGAAGCGTTTGTTTATCAGCTCTTCCACCATCCACCTCTCCCCTCCCCAGGTCCC 780

Qy     781 CGATCTAAAATCAAAGAGATTGATTTAGGATGGGTGGGTGCCTTGTCTTCTCTCATTGTT 840
Db     781 CGATCTAAAATCAAAGAGATTGATTTAGGATGGGTGGGTGCCTTGTCTTCTCTCATTGTT 840

Qy     841 CGACATTTTAGTTACGTTTTCTCTGAGCTCTCTGGAAAGCATAAAAGTATAATATCTGTT 900
Db     841 CGACATTTTAGTTACGTTTTCTCTGAGCTCTCTGGAAAGCATAAAAGTATAATATCTGTT 900

Qy     901 AAAAGTTGGATGAATGAACTAATGAACGCAATGGGATTCCAGAAAACTCTGCGGGAGATG 960
Db     901 AAAAGTTGGATGAATGAACTAATGAACGCAATGGGATTCCAGAAAACTCTGCGGGAGATG 960

Qy     961 GGCTAGAGGACGAGGAGGAGGTGGATGAATCAGCCATGTTAGAGAGCCTGGGAAGGTGAG 1020
Db     961 GGCTAGAGGACGAGGAGGAGGTGGATGAATCAGCCATGTTAGAGAGCCTGGGAAGGTGAG 1020

Qy    1021 CAGAGTTGAAAACTTGATAGATCTAATAATTTACTGGCTCTGGGTTTGTCAGTCACTACA 1080
Db    1021 CAGAGTTGAAAACTTGATAGATCTAATAATTTACTGGCTCTGGGTTTGTCAGTCACTACA 1080
```

```
Qy   1081 TTGCAGCAAATGAGATTAGAGCATAGTTGTGGGAGGGAAGGAGGTGACGCAGCAATCTAT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TTGCAGCAAATGAGATTAGAGCATAGTTGTGGGAGGGAAGGAGGTGACGCAGCAATCTAT 1140

Qy   1141 TTGCACCTAGAAATTTTAGGCAAGTGATAGCTGCGTAATCATACTGCGGCACCGTTTTTT 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 TTGCACCTAGAAATTTTAGGCAAGTGATAGCTGCGTAATCATACTGCGGCACCGTTTTTT 1200

Qy   1201 TCTTGCAGCAGTAGCTGCTTGCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 TCTTGCAGCAGTAGCTGCTTGCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCG 1260

Qy   1261 GCTCTCTCCTGCAGGATCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGTC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 GCTCTCTCCTGCAGGATCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGTC 1320

Qy   1321 GCCACTGCTTCAGTCCGGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTG 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 GCCACTGCTTCAGTCCGGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTG 1380

Qy   1381 GATCGTCGTTCCCTCAGTCCGGACGCCACTGCTAGGTCCGACCACCGCCGCTTCTGATAT 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 GATCGTCGTTCCCTCAGTCCGGACGCCACTGCTAGGTCCGACCACCGCCGCTTCTGATAT 1440

Qy   1441 TTCGGTGAGTCTTTTCCTGTGGAGGTTTGGTCTCCCGATCTCTGTGGTAGCCACCTTAGG 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 TTCGGTGAGTCTTTTCCTGTGGAGGTTTGGTCTCCCGATCTCTGTGGTAGCCACCTTAGG 1500

Qy   1501 CGTGTACGGTCCTTTGAAAA 1520
          ||||||||||||||||||||
Db   1501 CGTGTACGGTCCTTTGAAAA 1520
```

RESULT 3
ABK10802
ID ABK10802 standard; DNA; 2720 BP.
XX
AC ABK10802;
XX
DT 21-MAY-2002 (first entry)
XX
DE Human nucleosomal assembly protein IL2 (NAPIL2 or BPX) gene.
XX
KW Human nucleosomal assembly protein IL2; NAPIL2; BPX; nootropic;
KW neurotropic; cytostatic; neural system defect; cancer;
KW neuronal cell tumour; gene therapy; neurodegenerative disease;
KW Parkinson's disease; Alzheimer's disease; spina bifida; anencephaly;
KW brain structure; brain organisation; X-chromosome linked disorder;
KW nucleosome activity; cell cycle; cyclin; transcription factor control;
KW histone; gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 525..910
FT /*tag= a
FT /product= "NAPIL2"
FT /note= "Nucleosomal assembly protein IL2"
FT
XX
PN WO200185995-A2.
XX
PD 15-NOV-2001.
XX
PF 04-MAY-2001; 2001WO-IB00960.
XX
PR 05-NOV-2000; 2000US-202111P.
XX
PA (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Avner P, Rogner UC, Spyropoulos D, Rougeulle C;
XX
DR WPI; 2002-075254/10.
XX
PT Screening neural system defects in mammal, comprises detecting a
PT modification of NAPIL2 and/or NAPIL2 gene chromosomal material, that
PT causes a loss of biological function and correlating the modification
PT with the defect -
XX
PS Disclosure; Fig 9; 83pp; English.
XX
CC The invention describes a method of screening mammalian neural system
CC defects, comprising detecting a modification of the nucleosomal assembly
CC protein NAP1L2 and/or Napl12 gene in a human and/or mouse, where
CC modification is substitution, deletion, frame-shift, insertion aberrant
CC or altered epigenetic control, that causes a loss of function in the gene
CC and correlating the gene modification with a potential for a neural
CC system defect. The invention includes a neuronal cell containing a
CC recombinant chromosome on which Nap1L2 is located. The cell is useful for
CC screening therapeutic compounds by introducing a compound to be screened
CC and correlating change in the proliferation of cells with the activity of
CC the compound, where change in proliferating cells is preferably a control
CC of cancer of neural cells. The Nap1L2 promoter is useful for predicting
CC tumour of neural origin, and for genetic therapy of neurodegenerative
CC diseases such as Parkinson's, Alzheimer's diseases or accidents. Nap1L2
CC is also implicated in spina bifida with or without anencephaly, loss of
CC brain structure/organisation, X-chromosome linked disorders and
CC inappropriate control of nucleosome activity in neurons, neuronal cell
CC cycle, cyclins, DNA binding transcription factors, histones and histone
CC shuttling. This sequence represents the human nucleosomal assembly
CC protein IL2 (NAPIL2 or BPX) gene, a homologue of the murine nucleosomal
CC assembly protein IL2 (NapIL2 or Bpx) gene described in the invention.
XX
SQ Sequence 2720 BP; 808 A; 461 C; 645 G; 806 T; 0 other;

Query Match 34.4%; Score 522.4; DB 24; Length 2720;
Best Local Similarity 99.8%; Pred. No. 2.3e-110;
Matches 523; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy    997 TGTTAGAGAGCCTGGGAAGGTGAGCAGAGTTGAAAACTTGATAGATCTAATAATTTACTG 1056
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TGTTAGAGAGCCTGGGAAGGTGAGCAGAGCTGAAAACTTGATAGATCTAATAATTTACTG 60

Qy   1057 GCTCTGGGTTTGTCAGTCACTACATTGCAGCAAATGAGATTAGAGCATAGTTGTGGGAGG 1116
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCTCTGGGTTTGTCAGTCACTACATTGCAGCAAATGAGATTAGAGCATAGTTGTGGGAGG 120

Qy   1117 GAAGGAGGTGACGCAGCAATCTATTTGCACCTAGAAATTTTAGGCAAGTGATAGCTGCGT 1176
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GAAGGAGGTGACGCAGCAATCTATTTGCACCTAGAAATTTTAGGCAAGTGATAGCTGCGT 180

Qy   1177 AATCATACTGCGGCACCGTTTTTTTCTTGCAGCAGTAGCTGCTTGCGGAGGAGGTCTGCC 1236
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AATCATACTGCGGCACCGTTTTTTTCTTGCAGCAGTAGCTGCTTGCGGAGGAGGTCTGCC 240

Qy   1237 CACTGCAGCTCTCTGCAGTCTCCGGCTCTCTCCTGCAGGATCGGTCAACGCAGCCGTCGC 1296
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CACTGCAGCTCTCTGCAGTCTCCGGCTCTCTCCTGCAGGATCGGTCAACGCAGCCGTCGC 300

Qy   1297 CGCCCTCTGCACCCAGCCCAGGTCGCCACTGCTTCAGTCCGGTTCTCAAAGCCTCAGCAC 1356
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db.   301 CGCCCTCTGCACCCAGCCCAGGTCGCCACTGCTTCAGTCCGGTTCTCAAAGCCTCAGCAC 360

Qy   1357 CATCTTTTATCCCCGAGCAGCCTGGATCGTCGTTCCCTCAGTCCGGACGCCACTGCTAGG 1416
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CATCTTTTATCCCCGAGCAGCCTGGATCGTCGTTCCCTCAGTCCGGACGCCACTGCTAGG 420

Qy   1417 TCCGACCACCGCCGCTTCTGATATTTCGGTGAGTCTTTTCCTGTGGAGGTTTGGTCTCCC 1476
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TCCGACCACCGCCGCTTCTGATATTTCGGTGAGTCTTTTCCTGTGGAGGTTTGGTCTCCC 480

Qy   1477 GATCTCTGTGGTAGCCACCTTAGGCGTGTACGGTCCTTTGAAAA 1520
          |||||||||||||||||||||||||||||||||||||||||||||
Db    481 GATCTCTGTGGTAGCCACCTTAGGCGTGTACGGTCCTTTGAAAA 524
```

```
RESULT 4
AAC01166
ID   AAC01166 standard; cDNA; 629 BP.
XX
AC   AAC01166;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 1164.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000
XX
[redacted] 21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; AAG01160.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 1164; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. An ORF has been identified within the
CC   sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC   derived from 30 different tissues. EST sequences usually correspond
CC   mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC   often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC   well suited for isolating cDNA sequences derived from the 5' ends of
CC   mRNAs and even in those cases where longer cDNA sequences have been
CC   obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC   mRNAs with intact 5' ends and can therefore be used to obtain full length
CC   cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
[redacted] gene therapy and chromosome mapping procedures. They are used to obtain
CC   upstream regulatory sequences and to design expression and secretion
CC   vectors.
XX
SQ   Sequence 629 BP; 128 A; 162 C; 191 G; 143 T; 5 other;

  Query Match             25.0%;  Score 380.6;  DB 21;  Length 629;
  Best Local Similarity   99.7%;  Pred. No. 6.1e-78;
  Matches  380;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1140 TTTGCACCTAGAAATTTTAGGCAAGTGATAGCTGCGTAATCATACTGCGGCACCGTTTTT 1199
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3 TTTGCACCTAGAAATTTTAGGCAAGTGATAGCTGCGTAATCATACTGCGGCACCGTTTTT 62

Qy   1200 TTCTTGCAGCAGTAGCTGCTTGCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCC 1259
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     63 TTCTTGCAGCAGTAGCTGCTTGCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCC 122

Qy   1260 GGCTCTCTCCTGCAGGATCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGT 1319
           |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db    123 GGCTCTCTCCTGCAGGMTCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGT 182

Qy   1320 CGCCACTGCTTCAGTCCGGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCT 1379
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183 CGCCACTGCTTCAGTCCGGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCT 242

Qy   1380 GGATCGTCGTTCCCTCAGTCCGGACGCCACTGCTAGGTCCGACCACCGCCGCTTCTGATA 1439
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    243 GGATCGTCGTTCCCTCAGTCCGGACGCCACTGCTAGGTCCGACCACCGCCGCTTCTGATA 302

Qy   1440 TTTCGGTGAGTCTTTTCCTGTGGAGGTTTGGTCTCCCGATCTCTGTGGTAGCCACCTTAG 1499
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    303 TTTCGGTGAGTCTTTTCCTGTGGAGGTTTGGTCTCCCGATCTCTGTGGTAGCCACCTTAG 362

Qy   1500 GCGTGTACGGTCCTTTGAAAA 1520
           |||||||||||||||||||||
Db    363 GCGTGTACGGTCCTTTGAAAA 383


RESULT 5
ABK10360
ID   ABK10360 standard; DNA; 1725 BP.
XX
AC   ABK10360;
XX
DT   21-MAY-2002  (first entry)
XX
DE   Murine nucleosomal assembly protein IL2 (NapIL2 or Bpx) promoter.
XX
KW   Murine nucleosomal assembly protein IL2; NapIL2; Bpx; nootropic;
KW   neurotropic; cytostatic; neural system defect; cancer;
KW   neuronal cell tumour; gene therapy; neurodegenerative disease;
KW   Parkinson's disease; Alzheimer's disease; spina bifida; anencephaly;
KW   brain structure; brain organisation; X-chromosome linked disorder;
KW   nucleosome activity; cell cycle; cyclin; transcription factor control;
KW   histone; promoter; ds; mouse.
XX
OS   Mus sp.
XX
PN   WO200185995-A2.
XX
PD   15-NOV-2001.
XX
PF   04-MAY-2001; 2001WO-IB00960.
XX
PR   05-MAY-2000; 2000US-202111P.
XX
PA   (INSP ) INST PASTEUR.
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Avner P,  Rogner UC,  Spyropoulos D,  Rougeulle C;
XX
DR   WPI; 2002-075254/10.
XX
PT   Screening neural system defects in mammal, comprises detecting a
PT   modification of NAP1L2 and/or Nap112 gene chromosomal material, that
PT   causes a loss of biological function and correlating the modification
PT   with the defect -
XX
PS   Claim 32; Fig 7; 83pp; English.
XX
CC   The invention describes a method of screening mammalian neural system
CC   defects, comprising detecting a modification of the nucleosomal assembly
CC   protein NAP1L2 and/or Nap112 gene in a human and/or mouse, where
CC   modification is substitution, deletion, frame-shift, insertion aberrant
CC   or altered epigenetic control, that causes a loss of function in the gene
CC   and correlating the gene modification with a potential for a neural
CC   system defect. The invention includes a neuronal cell containing a
CC   recombinant chromosome on which Nap1L2 is located. The cell is useful for
CC   screening therapeutic compounds by introducing a compound to be screened
CC   and correlating change in the proliferation of cells with the activity of
CC   the compound, where change in proliferating cells is preferably a control
CC   of cancer of neural cells. The Nap1L2 promoter is useful for predicting
CC   tumour of neural origin, and for genetic therapy of neurodegenerative
CC   diseases such as Parkinson's, Alzheimer's diseases or accidents. Nap1L2
CC   is also implicated in spina bifida with or without anencephaly, loss of
CC   brain structure/organisation, X-chromosome linked disorders and
CC   inappropriate control of nucleosome activity in neurons, neuronal cell
CC   cycle, cyclins, DNA binding transcription factors, histones and histone
```

```
CC   shuttling. This sequence represents a murine nucleosomal assembly protein
CC   IL2 (NapIL2 or Bpx) promoter fragment, described in the invention.
XX
SQ   Sequence 1725 BP; 428 A; 443 C; 391 G; 463 T; 0 other;

  Query Match             15.4%;  Score 233.8;  DB 24;  Length 1725;
  Best Local Similarity   65.1%;  Pred. No. 5.4e-44;
  Matches  486;  Conservative    0;  Mismatches  222;  Indels   39;  Gaps    8;

Qy    580 CTCCCCTCACAGGAAATGTTTCGAGAATTTTTCAACCTAAAATCATATAGCTTGTGAAAA 639
Db     74 CTTCCTCATCAGGAAATATTATGAGAATTTTCCCATTTAAAATCACACAGGTTGTGAAAA 133

Qy    640 ATACCGACAAACATAATATAGAATATTTAAATAACTGAC----ACGCCACCTAAAGACCA 695
Db    134 TTACAGA-AACCAGGGTACAGAATATTTAAACCACTGTCAGTTACATCATCCAAAGGCCA 192

Qy    696 TCAGTGCTAATTCCTGGTGTTTTTAATCTTTGAAGCGTTTGTTTATCAGCTCTTCCACCA 755
Db    193 CCTATGCTTATTTTTGGTAATTTTAAACCTCAAAGGATCTCTTTGTGGGCTCCTCCACTA 252

Qy    756 TCCACCTCTCCCCTCCCCAGGTCCCCGATCTAAAATCAAAGAGATTGATTTAGGATGGGT 815
Db    253 CCCTCCTCT--CTTTCCCAGAGCCTCAGGTTATAACCAAAGGGATAGACTAAAGACAATC 310

Qy    816 GGGTGCCTTGTCTTCTCTCATTGTTCGACATTTTAGTTACGTTTTCTCTGAGCTCTCTGG 875
Db    311 CAGTACCTTGCCCATTTTTTTCATTC-----CTTGTCACTGTTTCCATATAGCTCTTTTG 365

Qy    876 AAAGCATAAAAGTATAATATCTGTTAAAAGTTGGATGAATGAACTAATGAACGCAATGGG 935
Db    366 AAATTATGAACATATAGTATCAGTTGAAAACGGAATGAATGA-----------TACTGCA 414

Qy    936 ATTCCAGAAAACTCTGCGGGAGATGGGCTAGAGGACGAGGAGGAGGTGGATGAATCAGCC 995
Db    415 TTTCTGCAAAATTCCACAGGCTATAGGGTGGAAGATGAGCCATAGGTGGAGGAATCAGCC 474

Qy    996 ATGTTAGAGAGCCTGGGAAGGTGAGCAGAGTTGAAAACTTGATAGATCTAATAATTTACT 1055
Db    475 ATATTAGAGAATCTGGGAAGGCAAGAGGTGTTGAAATTTTGATTCATCTACTAATTTACT 534

Qy   1056 GGCTCTGGGTTTGTCAGTCACTACATTGCAGCAAATGAGATTAGAGCATAGTTGTGGGAG 1115
Db    535 GGCTCAGGATTTGTCAATCACTGCAGCCTGGCAAATGAGATTAGAGAAGAGTCCTGGGAG 594

Qy   1116 GGAAGGAGGTGACGCAGCAATCTATTTGCACCT-AGAAATTTTAGGCAAGTGATAGCTGC 1174
Db    595 GGAAGG-GGTGACGCAGCAACCTGCATACACTTAAAAAAAAAGAGCTGAGAGACAACTGC 653

Qy   1175 GTAATCATACTGCGGCACCGTTTTTTTCTT--------------GCAGCAGTAGCTGCTT 1220
Db    654 GTAATCATACTGCGGCACCAGTTCCTCCATCCCTCCGCCCCCGAGTGGCTGGAGCAGCTG 713

Qy   1221 GCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCGGCTCTCTCCTGCAGGATCGG 1280
Db    714 CTTGCGGAGGTCTGCCCACTGCGGCTCTCTGCAGTCTCTAGCCTGTTCCTTCAGGGCCTA 773

Qy   1281 TCAACGCAGCCGTCGCCGCCCTCTGCA 1307
Db    774 GAGTCTCCGCCCAGACAGCCGGTTTCA 800
```

```
RESULT 6
ABK10801
ID   ABK10801 standard; DNA; 2819 BP.
XX
AC   ABK10801;
XX
DT   21-MAY-2002  (first entry)
XX
DE   Murine nucleosomal assembly protein IL2 (NapIL2 or Bpx) coding sequence.
XX
KW   Murine nucleosomal assembly protein I12; NapIL2; Bpx; nootropic;
KW   neurotropic; cytostatic; neural system defect; cancer;
KW   neuronal cell tumour; gene therapy; neurodegenerative disease;
KW   Parkinson's disease; Alzheimer's disease; spina bifida; anencephaly;
KW   brain structure; brain organisation; X-chromosome linked disorder;
KW   nucleosome activity; cell cycle; cyclin; transcription factor control;
KW   histone; gene; ds; mouse.
XX
OS   Mus sp.
XX
PN   WO200185995-A2.
XX
PD   15-NOV-2001.
XX
PF   04-MAY-2001; 2001WO-IB00960.
XX
PR   05-MAY-2000; 2000US-202111P.
XX
PA   (INSP ) INST PASTEUR.
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Avner P,  Rogner UC,  Spyropoulos D,  Rougeulle C;
XX
DR   WPI; 2002-075254/10.
XX
PT   Screening neural system defects in mammal, comprises detecting a
PT   modification of NapIL2 and/or NapIL2 gene chromosomal material, that
PT   causes a loss of biological function and correlating the modification
PT   with the defect  -
XX
PS   Disclosure; Fig 8; 83pp; English.
XX
CC   The invention describes a method of screening mammalian neural system
CC   defects, comprising detecting a modification of the nucleosomal assembly
CC   protein NAP1L2 and/or Nap1l2 gene in a human and/or mouse, where
CC   modification is substitution, deletion, frame-shift, insertion aberrant
CC   or altered epigenetic control, that causes a loss of function in the gene
CC   and correlating the gene modification with a potential for a neural
CC   system defect. The invention includes a neuronal cell containing a
CC   recombinant chromosome on which Nap1L2 is located. The cell is useful for
CC   screening therapeutic compounds by introducing a compound to be screened
CC   and correlating change in the proliferation of cells with the activity of
CC   the compound, where change in proliferating cells is preferably a control
CC   of cancer of neural cells. The Nap1L2 promoter is useful for predicting
CC   tumour of neural origin, and for genetic therapy of neurodegenerative
CC   diseases such as Parkinson's, Alzheimer's diseases or accidents. Nap1L2
CC   is also implicated in spina bifida with or without anencephaly, loss of
CC   brain structure/organisation, X-chromosome linked disorders and
CC   inappropriate control of nucleosome activity in neurons, neuronal cell
CC   cycle, cyclins, DNA binding transcription factors, histones and histone
CC   shuttling. This sequence represents the murine nucleosomal assembly
CC   protein IL2 (NapIL2 or Bpx) gene, described in the invention.
XX
SQ   Sequence 2819 BP; 827 A; 522 C; 670 G; 800 T; 0 other;

  Query Match             11.0%;  Score 166.8;  DB 24;  Length 2819;
  Best Local Similarity   64.9%;  Pred. No. 1.7e-28;
  Matches  322;  Conservative    0;  Mismatches  147;  Indels   27;  Gaps    4;

Qy    827 CTTCTCTCATTGTTCGACATTTTAGTTACGTTTTCTCTGAGCTCTCTGGAAAGCATAAAA 886
Db      5 CTTGCCCATTTTTTTCATTCCTTGTCACTGTTTCCATATAGCTCTTTTGAAATTATGAAC 64

Qy    887 GTATAATATCTGTTAAAAGTTGGATGAATGAACTAATGAACGCAATGGGATTCCAGAAAA 946
Db     65 ATATAGTATCAGTTGAAAACGGAATGAATGA-----------TACTGCATTTCTGCAAAA 113

Qy    947 CTCTGCGGGAGATGGGCTAGAGGACGAGGAGGAGGTGGATGAATCAGCCATGTTAGAGAG 1006
Db    114 TTCCACAGGCTATAGGGTGGAAGATGAGCCATAGGTGGAGGAATCAGCCATATTAGAGAA 173

Qy   1007 CCTGGGAAGGTGAGCAGAGTTGAAAACTTGATAGATCTAATAATTTACTGGCTCTGGGTT 1066
Db    174 TCTGGGAAGGCAAGAGGTGTTGAAATTTTGATTCATCTACTAATTTACTGGCTCAGGATT 233
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

Run on: January 8, 2003, 16:38:10 ; Search time 3929 Seconds
(without alignments)
11258.912 Million cell updates/sec

Title: US-09-847-665-4
Perfect score: 1520
Sequence: 1 acttaaaggaaaaatttatc.....cgtgtacggtcctttgaaaa 1520

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1520	100.0	1520	6	AX327802	AX327802 Sequence
2	1520	100.0	3699	6	AX327804	AX327804 Sequence
3	1520	100.0	153578	9	AC004074	AC004074 Homo Sapi
4	522.4	34.4	2720	6	AX327801	AX327801 Sequence
5	356	23.4	2515	9	AB027013	AB027013 Homo sapi
6	299.2	19.7	2524	9	BC026325	BC026325 Homo sapi
7	262	17.2	272	9	HS84H6R	Z63470 H.sapiens C
8	233.8	15.4	1725	6	AX327799	AX327799 Sequence
c 9	233.8	15.4	210614	2	AL773526	AL773526 Mus muscu
10	233.8	15.4	214384	10	AJ421480	AJ421480 Mus muscu
11	166.8	11.0	2819	6	AX327800	AX327800 Sequence
c 12	70.4	4.6	2226	3	MIDVTRN	X05914 D. virilis
c 13	69.8	4.6	154995	9	AC011979	AC011979 Homo sapi
14	69.6	4.6	152409	2	PFMAL1P1	AL031744 Plasmodiu
c 15	69.4	4.6	108902	2	AC011430	AC011430 Homo sapi
16	68.6	4.5	123589	9	AC104790	AC104790 Homo sapi
17	68.6	4.5	172758	2	AC022553	AC022553 Homo sapi
18	68.4	4.5	224448	2	PFMAL4P4	AL035477 Plasmodiu
c 19	68.4	4.5	349980	6	AX344563	AX344563 Sequence
c 20	68.2	4.5	147192	2	AC116925	AC116925 Dictyoste
c 21	68.2	4.5	245802	2	AC006279	AC006279 Plasmodiu
22	68	4.5	118642	9	AC126283	AC126283 Homo sapi
23	68	4.5	159712	2	AP000792	AP000792 Homo sapi
24	68	4.5	193227	9	AP003499	AP003499 Homo sapi
25	67.4	4.4	2426	8	SDU49822	U49822 Saccharomyc
c 26	67.4	4.4	349980	6	AX344555	AX344555 Sequence
27	67.2	4.4	169841	2	AC073888	AC073888 Homo sapi
28	67.2	4.4	177476	2	AC103159	AC103159 Rattus no
29	67.2	4.4	199200	9	AL359633	AL359633 Human DNA
30	67	4.4	213535	6	AX197417	AX197417 Sequence
31	67	4.4	213535	6	AX223856	AX223856 Sequence
32	66.6	4.4	97683	2	AC116548	AC116548 Dictyoste
c 33	66.4	4.4	143331	9	AC091214	AC091214 Homo sapi
c 34	66	4.3	86827	3	PFMAL3P5	AL034556 Plasmodiu
35	65.4	4.3	10347	3	PFAVAR	L42244 Plasmodium
c 36	65.4	4.3	105686	9	AC012072	AC012072 Homo sapi
37	65.4	4.3	168698	9	AC068138	AC068138 Homo sapi
38	65.4	4.3	170796	2	AC069416	AC069416 Homo sapi
c 39	65.2	4.3	17967	6	AX345917	AX345917 Sequence
c 40	65.2	4.3	148202	2	AL772396	AL772396 Danio rer
41	65.2	4.3	209856	2	AC117020	AC117020 Rattus no
c 42	64.8	4.3	6419	6	AX345169	AX345169 Sequence
c 43	64.8	4.3	163964	2	AC106154	AC106154 Rattus no
c 44	64.8	4.3	206059	2	AC127383	AC127383 Homo sapi
c 45	64.6	4.2	103344	9	HS1100E15	AL035551 Human DNA

ALIGNMENTS

RESULT 1
AX327802
LOCUS AX327802 1520 bp DNA linear PAT 07-JAN-2002
DEFINITION Sequence 4 from Patent WO0185995.
ACCESSION AX327802
VERSION AX327802.1 GI:18098075
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Avner,P., Rogner,U.C., Spyropoulos,D. and Rougeulle,C.
TITLE Identification of neural defects associated with the nucleosomal assembly protein 112 gene

```
  JOURNAL   Patent: WO 0185995-A 4 15-NOV-2001;
            INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE
            SCIENTIFIQUE (CNRS) (FR)
FEATURES             Location/Qualifiers
     source          1. .1520
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      444 a    321 c    302 g    453 t
ORIGIN

  Query Match           100.0%;  Score 1520;  DB 6;  Length 1520;
  Best Local Similarity  100.0%;  Pred. No. 4e-312;
  Matches 1520;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ACTTAAAGGAAAAATTTATCTATAAACTGACAGAATTTAGAAATAAATACAACAATATGT 60
Db       1 ACTTAAAGGAAAAATTTATCTATAAACTGACAGAATTTAGAAATAAATACAACAATATGT 60

Qy      61 AAACAGTTTTAATATCTGTGATAGTAACAAATTCTTTAAATCTGGAAAATAATAGTCACT 120
Db      61 AAACAGTTTTAATATCTGTGATAGTAACAAATTCTTTAAATCTGGAAAATAATAGTCACT 120

Qy     121 TAAAATTTTAAAAAATTGTTCAATTAATAAATGATCCAAGTTAGAAATATGAACAAAATA 180
Db     121 TAAAATTTTAAAAAATTGTTCAATTAATAAATGATCCAAGTTAGAAATATGAACAAAATA 180

Qy     181 AACCTCACCAATAATTACTATAGAGAGGAAATTTTAATTACTGCAAAGCTTTCCATCCTA 240
Db     181 AACCTCACCAATAATTACTATAGAGAGGAAATTTTAATTACTGCAAAGCTTTCCATCCTA 240

Qy     241 TAAATACATTATCAAATAGTTTAACCATTTCTTTAATGCTGAGATTTAGATTATTTCCAA 300
Db     241 TAAATACATTATCAAATAGTTTAACCATTTCTTTAATGCTGAGATTTAGATTATTTCCAA 300

Qy     301 TTAACTCAAAAGCATCAAGCAAATGTTATGATTTCTAAGAATAAACATAACTTTCCATTT 360
Db     301 TTAACTCAAAAGCATCAAGCAAATGTTATGATTTCTAAGAATAAACATAACTTTCCATTT 360

Qy     361 TGGCTTTTGTATATATGTATATTTCTAACGGCTGTTAAAGCCAGCATTAAGAAGGAGAAG 420
Db     361 TGGCTTTTGTATATATGTATATTTCTAACGGCTGTTAAAGCCAGCATTAAGAAGGAGAAG 420

Qy     421 CAGAAAGTCAGTATTGGGACTGGGGTTATTTATAAGCCAGGCAACTGGTTAATTGTGGTT 480
Db     421 CAGAAAGTCAGTATTGGGACTGGGGTTATTTATAAGCCAGGCAACTGGTTAATTGTGGTT 480

Qy     481 AATTGTCTGGTATGTTTACTAGTCACGTAGTTGTATACACCATACTAGTTTTTCATCACA 540
Db     481 AATTGTCTGGTATGTTTACTAGTCACGTAGTTGTATACACCATACTAGTTTTTCATCACA 540

Qy     541 GGCCCTCATTCGCCCCCACTGCCATCGGACTTCCTCCTCCTCCCCTCACAGGAAATGTTT 600
Db     541 GGCCCTCATTCGCCCCCACTGCCATCGGACTTCCTCCTCCTCCCCTCACAGGAAATGTTT 600

Qy     601 CGAGAATTTTTCAACCTAAAATCATATAGCTTGTGAAAAATACCGACAAACATAATATAG 660
Db     601 CGAGAATTTTTCAACCTAAAATCATATAGCTTGTGAAAAATACCGACAAACATAATATAG 660

Qy     661 AATATTTAAATAACTGACACGCCACCTAAAGACCATCAGTGCTAATTCCTGGTGTTTTTA 720
Db     661 AATATTTAAATAACTGACACGCCACCTAAAGACCATCAGTGCTAATTCCTGGTGTTTTTA 720

Qy     721 ATCTTTGAAGCGTTTGTTTATCAGCTCTTCCACCATCCACCTCTCCCCTCCCCAGGTCCC 780
Db     721 ATCTTTGAAGCGTTTGTTTATCAGCTCTTCCACCATCCACCTCTCCCCTCCCCAGGTCCC 780

Qy     781 CGATCTAAAATCAAAGAGATTGATTTAGGATGGGTGGGTGCCTTGTCTTCTCTCATTGTT 840
Db     781 CGATCTAAAATCAAAGAGATTGATTTAGGATGGGTGGGTGCCTTGTCTTCTCTCATTGTT 840

Qy     841 CGACATTTTAGTTACGTTTTCTCTGAGCTCTCTGGAAAGCATAAAAGTATAATATCTGTT 900
Db     841 CGACATTTTAGTTACGTTTTCTCTGAGCTCTCTGGAAAGCATAAAAGTATAATATCTGTT 900

Qy     901 AAAAGTTGGATGAATGAACTAATGAACGCAATGGGATTCCAGAAAACTCTGCGGGAGATG 960
Db     901 AAAAGTTGGATGAATGAACTAATGAACGCAATGGGATTCCAGAAAACTCTGCGGGAGATG 960

Qy     961 GGCTAGAGGACGAGGAGGAGGTGGATGAATCAGCCATGTTAGAGAGCCTGGGAAGGTGAG 1020
Db     961 GGCTAGAGGACGAGGAGGAGGTGGATGAATCAGCCATGTTAGAGAGCCTGGGAAGGTGAG 1020

Qy    1021 CAGAGTTGAAAACTTGATAGATCTAATAATTTACTGGCTCTGGGTTTGTCAGTCACTACA 1080
Db    1021 CAGAGTTGAAAACTTGATAGATCTAATAATTTACTGGCTCTGGGTTTGTCAGTCACTACA 1080

Qy    1081 TTGCAGCAAATGAGATTAGAGCATAGTTGTGGGAGGGAAGGAGGTGACGCAGCAATCTAT 1140
Db    1081 TTGCAGCAAATGAGATTAGAGCATAGTTGTGGGAGGGAAGGAGGTGACGCAGCAATCTAT 1140

Qy    1141 TTGCACCTAGAAATTTTAGGCAAGTGATAGCTGCGTAATCATACTGCGGCACCGTTTTTT 1200
Db    1141 TTGCACCTAGAAATTTTAGGCAAGTGATAGCTGCGTAATCATACTGCGGCACCGTTTTTT 1200

Qy    1201 TCTTGCAGCAGTAGCTGCTTGCGGAGGAGGTCTGCCCCACTGCAGCTCTCTGCAGTCTCCG 1260
Db    1201 TCTTGCAGCAGTAGCTGCTTGCGGAGGAGGTCTGCCCCACTGCAGCTCTCTGCAGTCTCCG 1260

Qy    1261 GCTCTCTCCTGCAGGATCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGTC 1320
Db    1261 GCTCTCTCCTGCAGGATCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGTC.1320

Qy    1321 GCCACTGCTTCAGTCCGGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTG 1380
Db    1321 GCCACTGCTTCAGTCCGGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTG 1380

Qy    1381 GATCGTCGTTCCCTCAGTCCGGACGCCACTGCTAGGTCCGACCACCGCCGCTTCTGATAT 1440
Db    1381 GATCGTCGTTCCCTCAGTCCGGACGCCACTGCTAGGTCCGACCACCGCCGCTTCTGATAT 1440

Qy    1441 TTCGGTGAGTCTTTTCCTGTGGAGGTTTGGTCTCCCGATCTCTGTGGTAGCCACCTTAGG 1500
Db    1441 TTCGGTGAGTCTTTTCCTGTGGAGGTTTGGTCTCCCGATCTCTGTGGTAGCCACCTTAGG 1500

Qy    1501 CGTGTACGGTCCTTTGAAAA 1520
Db    1501 CGTGTACGGTCCTTTGAAAA 1520

RESULT 2
AX327804
LOCUS       AX327804                3699 bp    DNA     linear   PAT 07-JAN-200
DEFINITION  Sequence 6 from Patent WO0185995.
ACCESSION   AX327804
VERSION     AX327804.1  GI:18098077
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Avner,P., Rogner,U.C., Spyropoulos,D. and Rougeulle,C.
  TITLE     Identification of neural defects associated with the nucleosomal
            assembly protein 112 gene
  JOURNAL   Patent: WO 0185995-A 6 15-NOV-2001;
            INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE
            SCIENTIFIQUE (CNRS) (FR)
FEATURES             Location/Qualifiers
     source          1. .3699
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      1144 a    636 c    805 g    1114 t
ORIGIN

  Query Match           100.0%;  Score 1520;  DB 6;  Length 3699;
  Best Local Similarity  100.0%;  Pred. No. 4.1e-312;
```

```
Matches 1520;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ACTTAAAGGAAAAATTTATCTATAAACTGACAGAATTTAGAAATAAATACAACAATATGT 60
Db       1 ACTTAAAGGAAAAATTTATCTATAAACTGACAGAATTTAGAAATAAATACAACAATATGT 60

Qy      61 AAACAGTTTTAATATCTGTGATAGTAACAAATTCTTTAAATCTGGAAAATAATAGTCACT 120
Db      61 AAACAGTTTTAATATCTGTGATAGTAACAAATTCTTTAAATCTGGAAAATAATAGTCACT 120

Qy     121 TAAAATTTTAAAAAAATTGTTCAATTAATAAATGATCCAAGTTAGAAATATGAACAAAATA 180
Db     121 TAAAATTTTAAAAAAATTGTTCAATTAATAAATGATCCAAGTTAGAAATATGAACAAAATA 180

Qy     181 AACCTCACCAATAATTACTATAGAGAGGAAATTTTAATTACTGCAAAGCTTTCCATCCTA 240
Db     181 AACCTCACCAATAATTACTATAGAGAGGAAATTTTAATTACTGCAAAGCTTTCCATCCTA 240

Qy     241 TAAATACATTATCAAATAGTTTAACCATTTCTTTAATGCTGAGATTTAGATTATTTCCAA 300
I      241 TAAATACATTATCAAATAGTTTAACCATTTCTTTAATGCTGAGATTTAGATTATTTCCAA 300

Qy     301 TTAACTCAAAAGCATCAAGCAAATGTTATGATTTCTAAGAATAAACATAACTTTCCATTT 360
Db     301 TTAACTCAAAAGCATCAAGCAAATGTTATGATTTCTAAGAATAAACATAACTTTCCATTT 360

Qy     361 TGGCTTTTGTATATATGTATATTTCTAACGGCTGTTAAAGCCAGCATTAAGAAGGAGAAG 420
Db     361 TGGCTTTTGTATATATGTATATTTCTAACGGCTGTTAAAGCCAGCATTAAGAAGGAGAAG 420

Qy     421 CAGAAAGTCAGTATTGGGACTGGGGTTATTTATAAGCCAGGCAACTGGTTAATTGTGGTT 480
Db     421 CAGAAAGTCAGTATTGGGACTGGGGTTATTTATAAGCCAGGCAACTGGTTAATTGTGGTT 480

Qy     481 AATTGTCTGGTATGTTTACTAGTCACGTAGTTGTATACACCATACTAGTTTTTCATCACA 540
Db     481 AATTGTCTGGTATGTTTACTAGTCACGTAGTTGTATACACCATACTAGTTTTTCATCACA 540

Qy     541 GGCCCTCATTCGCCCCCACTGCCATCGGACTTCCTCCTCCTCCCCTCACAGGAAATGTTT 600
Db     541 GGCCCTCATTCGCCCCCACTGCCATCGGACTTCCTCCTCCTCCCCTCACAGGAAATGTTT 600

Qy     601 CGAGAATTTTTCAACCTAAAATCATATAGCTTGTGAAAAATACCGACAAACATAATATAG 660
Db     601 CGAGAATTTTTCAACCTAAAATCATATAGCTTGTGAAAAATACCGACAAACATAATATAG 660

Qy     661 AATATTTAAATAACTGACACGCCACCTAAAGACCATCAGTGCTAATTCCTGGTGTTTTTA 720
I      661 AATATTTAAATAACTGACACGCCACCTAAAGACCATCAGTGCTAATTCCTGGTGTTTTTA 720

Qy     721 ATCTTTGAAGCGTTTGTTTATCAGCTCTTCCACCATCCACCTCTCCCCTCCCCAGGTCCC 780
Db     721 ATCTTTGAAGCGTTTGTTTATCAGCTCTTCCACCATCCACCTCTCCCCTCCCCAGGTCCC 780

Qy     781 CGATCTAAAATCAAAGAGATTGATTTAGGATGGGTGGGTGCCTTGTCTTCTCTCATTGTT 840
Db     781 CGATCTAAAATCAAAGAGATTGATTTAGGATGGGTGGGTGCCTTGTCTTCTCTCATTGTT 840

Qy     841 CGACATTTTAGTTACGTTTTCTCTGAGCTCTCTGGAAAGCATAAAAGTATAATATCTGTT 900
Db     841 CGACATTTTAGTTACGTTTTCTCTGAGCTCTCTGGAAAGCATAAAAGTATAATATCTGTT 900

Qy     901 AAAAGTTGGATGAATGAACTAATGAACGCAATGGGATTCCAGAAAACTCTGCGGGAGATG 960
Db     901 AAAAGTTGGATGAATGAACTAATGAACGCAATGGGATTCCAGAAAACTCTGCGGGAGATG 960

Qy     961 GGCTAGAGGACGAGGAGGAGGTGGATGAATCAGCCATGTTAGAGAGCCTGGGAAGGTGAG 1020
Db     961 GGCTAGAGGACGAGGAGGAGGTGGATGAATCAGCCATGTTAGAGAGCCTGGGAAGGTGAG 1020

Qy    1021 CAGAGTTGAAAACTTGATAGATCTAATAATTTACTGGCTCTGGGTTTGTCAGTCACTACA 1080
Db    1021 CAGAGTTGAAAACTTGATAGATCTAATAATTTACTGGCTCTGGGTTTGTCAGTCACTACA 1080

Qy    1081 TTGCAGCAAATGAGATTAGAGCATAGTTGTGGGAGGGAAGGAGGTGACGCAGCAATCTAT 1140
Db    1081 TTGCAGCAAATGAGATTAGAGCATAGTTGTGGGAGGGAAGGAGGTGACGCAGCAATCTAT 1140

Qy    1141 TTGCACCTAGAAATTTTAGGCAAGTGATAGCTGCGTAATCATACTGCGGCACCGTTTTTT 1200
Db    1141 TTGCACCTAGAAATTTTAGGCAAGTGATAGCTGCGTAATCATACTGCGGCACCGTTTTTT 1200

Qy    1201 TCTTGCAGCAGTAGCTGCTTGCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCG 1260
Db    1201 TCTTGCAGCAGTAGCTGCTTGCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCG 1260

Qy    1261 GCTCTCTCCTGCAGGATCGGTCAACGCAGCCGTCGCCGCCCCTCTGCACCCAGCCCAGGTC 1320
Db    1261 GCTCTCTCCTGCAGGATCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGTC 1320

Qy    1321 GCCACTGCTTCAGTCCGGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTG 1380
Db    1321 GCCACTGCTTCAGTCCGGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTG 1380

Qy    1381 GATCGTCGTTCCCTCAGTCCGGACGCCACTGCTAGGTCCGACCACCGCCGCTTCTGATAT 1440
Db    1381 GATCGTCGTTCCCTCAGTCCGGACGCCACTGCTAGGTCCGACCACCGCCGCTTCTGATAT 1440

Qy    1441 TTCGGTGAGTCTTTTCCTGTGGAGGTTTGGTCTCCCGATCTCTGTGGTAGCCACCTTAGG 1500
Db    1441 TTCGGTGAGTCTTTTCCTGTGGAGGTTTGGTCTCCCGATCTCTGTGGTAGCCACCTTAGG 1500

Qy    1501 CGTGTACGGTCCTTTGAAAA 1520
Db    1501 CGTGTACGGTCCTTTGAAAA 1520
```

RESULT 3
AC004074
LOCUS AC004074 153578 bp DNA linear PRI 12-APR-1998
DEFINITION Homo Sapiens Chromosome X clone bWXD759, complete sequence.
ACCESSION AC004074
VERSION AC004074.1 GI:3046270
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 153578)
AUTHORS Chen,E., Brownstein,B.H., States,D.J., Schlessinger,D. and Mazzarella,R.
TITLE Direct Submission
JOURNAL Unpublished (1997)
REFERENCE 2 (bases 1 to 153578)
AUTHORS Brownstein,B.H., States,D.J. and Mazzarella,R.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-1998) Center for Genetics in Medicine, Box 8232, Washington University School of Medicine, 4566 Scott Avenue, St. Louis, MO 63110, USA
REFERENCE 3 (bases 1 to 153578)
AUTHORS Brownstein,B.H., States,D.J. and Mazzarella,R.
TITLE Direct Submission
JOURNAL Submitted (12-APR-1998) Center for Genetics in Medicine, Box 8232, Washington University School of Medicine, 4566 Scott Avenue, St. Louis, MO 63110, USA
COMMENT On Apr 13, 1998 this sequence version replaced gi:2822135.
Current status of this project is available at:
'http://www.ibc.wustl.edu/cgm/seq_projects.html'
Submitted by:
Ellson Chen,
Advanced Center for Genetic Technology,
Applied Biosystems Division of Perlin Elmer Corp.,
850 Lincoln Center Drive,
Foster City, CA 94404 USA
e-mail: ellson@genseq.apldbio.com

and

Buddy Brownstein,
Center for Genetics in Medicine,
Washington University School of Medicine, Box 8232
4566 Scott Avenue,
St. Louis, MO 63110, USA
e-mail: buddy@genetics.wustl.edu

and

David J. States,
Institute for Biomedical Computing
Washington University in St. Louis
700 South Euclid Ave.
St. Louis, MO 63108 USA
e-mail: states@ibc.wustl.edu.

FEATURES Location/Qualifiers
source 1. .153578
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="bWXD759"

BASE COUNT 45069 a 29281 c 29096 g 50132 t
ORIGIN

Query Match 100.0%; Score 1520; DB 9; Length 153578;
Best Local Similarity 100.0%; Pred. No. 5.4e-312;
Matches 1520; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ACTTAAAGGAAAAATTTATCTATAAACTGACAGAATTTAGAAATAAATACAACAATATGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136351 ACTTAAAGGAAAAATTTATCTATAAACTGACAGAATTTAGAAATAAATACAACAATATGT 136410

Qy      61 AAACAGTTTTAATATCTGTGATAGTAACAAATTCTTTAAATCTGGAAAATAATAGTCACT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136411 AAACAGTTTTAATATCTGTGATAGTAACAAATTCTTTAAATCTGGAAAATAATAGTCACT 136470

Qy     121 TAAAATTTTAAAAAATTGTTCAATTAATAAATGATCCAAGTTAGAAATATGAACAAAATA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136471 TAAAATTTTAAAAAATTGTTCAATTAATAAATGATCCAAGTTAGAAATATGAACAAAATA 136530

Qy     181 AACCTCACCAATAATTACTATAGAGAGGAAATTTTAATTACTGCAAAGCTTTCCATCCTA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136531 AACCTCACCAATAATTACTATAGAGAGGAAATTTTAATTACTGCAAAGCTTTCCATCCTA 136590

Qy     241 TAAATACATTATCAAATAGTTTAACCATTTCTTTAATGCTGAGATTTAGATTATTTCCAA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136591 TAAATACATTATCAAATAGTTTAACCATTTCTTTAATGCTGAGATTTAGATTATTTCCAA 136650

Qy     301 TTAACTCAAAAGCATCAAGCAAATGTTATGATTTCTAAGAATAAACATAACTTTCCATTT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136651 TTAACTCAAAAGCATCAAGCAAATGTTATGATTTCTAAGAATAAACATAACTTTCCATTT 136710

Qy     361 TGGCTTTTGTATATATGTATATTTCTAACGGCTGTTAAAGCCAGCATTAAGAAGGAGAAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136711 TGGCTTTTGTATATATGTATATTTCTAACGGCTGTTAAAGCCAGCATTAAGAAGGAGAAG 136770

Qy     421 CAGAAAGTCAGTATTGGGACTGGGGTTATTTATAAGCCAGGCAACTGGTTAATTGTGGTT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136771 CAGAAAGTCAGTATTGGGACTGGGGTTATTTATAAGCCAGGCAACTGGTTAATTGTGGTT 136830

Qy     481 AATTGTCTGGTATGTTTACTAGTCACGTAGTTGTATACACCATACTAGTTTTTCATCACA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136831 AATTGTCTGGTATGTTTACTAGTCACGTAGTTGTATACACCATACTAGTTTTTCATCACA 136890

Qy     541 GGCCCTCATTCGCCCCCACTGCCATCGGACTTCCTCCTCCTCCCCTCACAGGAAATGTTT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136891 GGCCCTCATTCGCCCCCACTGCCATCGGACTTCCTCCTCCTCCCCTCACAGGAAATGTTT 136950

Qy     601 CGAGAATTTTTCAACCTAAAATCATATAGCTTGTGAAAAATACCGACAAACATAATATAG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136951 CGAGAATTTTTCAACCTAAAATCATATAGCTTGTGAAAAATACCGACAAACATAATATAG 137010

Qy     661 AATATTTAAATAACTGACACGCCACCTAAAGACCATCAGTGCTAATTCCTGGTGTTTTTA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137011 AATATTTAAATAACTGACACGCCACCTAAAGACCATCAGTGCTAATTCCTGGTGTTTTTA 137070

Qy     721 ATCTTTGAAGCGTTTGTTTATCAGCTCTTCCACCATCCACCTCTCCCCTCCCCAGGTCCC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137071 ATCTTTGAAGCGTTTGTTTATCAGCTCTTCCACCATCCACCTCTCCCCTCCCCAGGTCCC 137130

Qy     781 CGATCTAAAATCAAAGAGATTGATTTAGGATGGGTGGGTGCCTTGTCTTCTCTCATTGTT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137131 CGATCTAAAATCAAAGAGATTGATTTAGGATGGGTGGGTGCCTTGTCTTCTCTCATTGTT 137190

Qy     841 CGACATTTTAGTTACGTTTTCTCTGAGCTCTCTGGAAAGCATAAAAGTATAATATCTGTT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137191 CGACATTTTAGTTACGTTTTCTCTGAGCTCTCTGGAAAGCATAAAAGTATAATATCTGTT 137250

Qy     901 AAAAGTTGGATGAATGAACTAATGAACGCAATGGGATTCCAGAAAACTCTGCGGGAGATG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137251 AAAAGTTGGATGAATGAACTAATGAACGCAATGGGATTCCAGAAAACTCTGCGGGAGATG 137310

Qy     961 GGCTAGAGGACGAGGAGGAGGTGGATGAATCAGCCATGTTAGAGAGCCTGGGAAGGTGAG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137311 GGCTAGAGGACGAGGAGGAGGTGGATGAATCAGCCATGTTAGAGAGCCTGGGAAGGTGAG 137370

Qy    1021 CAGAGTTGAAAACTTGATAGATCTAATAATTTACTGGCTCTGGGTTTGTCAGTCACTACA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137371 CAGAGTTGAAAACTTGATAGATCTAATAATTTACTGGCTCTGGGTTTGTCAGTCACTACA 137430

Qy    1081 TTGCAGCAAATGAGATTAGAGCATAGTTGTGGGAGGGAAGGAGGTGACGCAGCAATCTAT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137431 TTGCAGCAAATGAGATTAGAGCATAGTTGTGGGAGGGAAGGAGGTGACGCAGCAATCTAT 137490

Qy    1141 TTGCACCTAGAAATTTTAGGCAAGTGATAGCTGCGTAATCATACTGCGGCACCGTTTTTT 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137491 TTGCACCTAGAAATTTTAGGCAAGTGATAGCTGCGTAATCATACTGCGGCACCGTTTTTT 137550

Qy    1201 TCTTGCAGCAGTAGCTGCTTGCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137551 TCTTGCAGCAGTAGCTGCTTGCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCG 137610

Qy    1261 GCTCTCTCCTGCAGGATCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGTC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137611 GCTCTCTCCTGCAGGATCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGTC 137670

Qy    1321 GCCACTGCTTCAGTCCGGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTG 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137671 GCCACTGCTTCAGTCCGGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTG 137730

Qy    1381 GATCGTCGTTCCCTCAGTCCGGACGCCACTGCTAGGTCCGACCACCGCCGCTTCTGATAT 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137731 GATCGTCGTTCCCTCAGTCCGGACGCCACTGCTAGGTCCGACCACCGCCGCTTCTGATAT 137790

Qy    1441 TTCGGTGAGTCTTTTCCTGTGGAGGTTTGGTCTCCCGATCTCTGTGGTAGCCACCTTAGG 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137791 TTCGGTGAGTCTTTTCCTGTGGAGGTTTGGTCTCCCGATCTCTGTGGTAGCCACCTTAGG 137850

Qy    1501 CGTGTACGGTCCTTTGAAAA 1520
          ||||||||||||||||||||
Db  137851 CGTGTACGGTCCTTTGAAAA 137870
```

RESULT 4
AX327801
LOCUS AX327801 2720 bp DNA linear PAT 07-JAN-2002
DEFINITION Sequence 3 from Patent WO0185995.
ACCESSION AX327801
VERSION AX327801.1 GI:18098074
KEYWORDS .
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```
REFERENCE   1
  AUTHORS   Avner,P., Rogner,U.C., Spyropoulos,D. and Rougeulle,C.
  TITLE     Identification of neural defects associated with the nucleosomal
            assembly protein 112 gene
  JOURNAL   Patent: WO 0185995-A 3 15-NOV-2001;
            INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE
            SCIENTIFIQUE (CNRS) (FR)
FEATURES             Location/Qualifiers
     source          1. .2720
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT       808 a     461 c     645 g     806 t
ORIGIN

  Query Match             34.4%;  Score 522.4;  DB 6;  Length 2720;
  Best Local Similarity   99.8%;  Pred. No. 1.7e-100;
  Matches  523;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    997 TGTTAGAGAGCCTGGGAAGGTGAGCAGAGTTGAAAACTTGATAGATCTAATAATTTACTG 1056
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      1 TGTTAGAGAGCCTGGGAAGGTGAGCAGAGCTGAAAACTTGATAGATCTAATAATTTACTG 60

Qy   1057 GCTCTGGGTTTGTCAGTCACTACATTGCAGCAAATGAGATTAGAGCATAGTTGTGGGAGG 1116
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCTCTGGGTTTGTCAGTCACTACATTGCAGCAAATGAGATTAGAGCATAGTTGTGGGAGG 120

Qy   1117 GAAGGAGGTGACGCAGCAATCTATTTGCACCTAGAAATTTTAGGCAAGTGATAGCTGCGT 1176
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GAAGGAGGTGACGCAGCAATCTATTTGCACCTAGAAATTTTAGGCAAGTGATAGCTGCGT 180

Qy   1177 AATCATACTGCGGCACCGTTTTTTTCTTGCAGCAGTAGCTGCTTGCGGAGGAGGTCTGCC 1236
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AATCATACTGCGGCACCGTTTTTTTCTTGCAGCAGTAGCTGCTTGCGGAGGAGGTCTGCC 240

Qy   1237 CACTGCAGCTCTCTGCAGTCTCCGGCTCTCTCCTGCAGGATCGGTCAACGCAGCCGTCGC 1296
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CACTGCAGCTCTCTGCAGTCTCCGGCTCTCTCCTGCAGGATCGGTCAACGCAGCCGTCGC 300

Qy   1297 CGCCCTCTGCACCCAGCCCAGGTCGCCACTGCTTCAGTCCGGTTCTCAAAGCCTCAGCAC 1356
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CGCCCTCTGCACCCAGCCCAGGTCGCCACTGCTTCAGTCCGGTTCTCAAAGCCTCAGCAC 360

Qy   1357 CATCTTTTATCCCCGAGCAGCCTGGATCGTCGTTCCCTCAGTCCGGACGCCACTGCTAGG 1416
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CATCTTTTATCCCCGAGCAGCCTGGATCGTCGTTCCCTCAGTCCGGACGCCACTGCTAGG 420

Qy   1417 TCCGACCACCGCCGCTTCTGATATTTCGGTGAGTCTTTTCCTGTGGAGGTTTGGTCTCCC 1476
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TCCGACCACCGCCGCTTCTGATATTTCGGTGAGTCTTTTCCTGTGGAGGTTTGGTCTCCC 480

Qy   1477 GATCTCTGTGGTAGCCACCTTAGGCGTGTACGGTCCTTTGAAAA 1520
          |||||||||||||||||||||||||||||||||||||||||||||
Db    481 GATCTCTGTGGTAGCCACCTTAGGCGTGTACGGTCCTTTGAAAA 524


RESULT 5
AB027013
LOCUS       AB027013                2515 bp    mRNA    linear   PRI 29-SEP-1999
DEFINITION  Homo sapiens mRNA for Nucleosome Assembly Protein 1-like 2,
            complete cds.
ACCESSION   AB027013
VERSION     AB027013.1  GI:5931609
KEYWORDS    Nucleosome Assembly Protein 1-like 2.
SOURCE      Homo sapiens cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2515)
  AUTHORS   Seki,N., Hattori,A., Hayashi,A., Kozuma,S., Muramatsu,M.,
            Miyajima,N. and Saito,T.
  TITLE     Nucleosome Assembly Protein 1-like 2 (Brain-Specific Protein,
            Xlinked)
  JOURNAL   Published Only in DataBase (1999)

REFERENCE   2  (bases 1 to 2515)
  AUTHORS   Seki,N., Hattori,A., Hayashi,A., Kozuma,S., Muramatsu,M.,
            Miyajima,N. and Saito,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-1999) Toshiyuki Saito, National Institute of
            Radiological Sciences, Genome Research Group; Anagawa 4-9-1,
            Inage-ku, Chiba 263-8555, Japan (E-mail:t_saito@nirs.go.jp,
            Tel:81-43-206-3135, Fax:81-43-251-9818)
FEATURES             Location/Qualifiers
     source          1. .2515
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             357. .1739
                     /note="Brain-Specific Protein, X linked"
                     /codon_start=1
                     /product="Nucleosome Assembly Protein 1-like 2"
                     /protein_id="BAA84706.1"
                     /db_xref="GI:5931610"
                     /translation="MAESENRKELSESSQEEAGNQIMVEGLGEHLERGEDAAAGLGDD
                     GKCGEEAAAGLGEEGENGEDTAAGSGEDGKKGGDTDEDSEADRPKGLIGYVLDTDPVE
                     SLPVKVKYRVLALKKLQTRAANLESKFLREFHDIERKFAEMYQPLLEKRRQIINAIYE
                     PTEEECEYKSDSEDCDDEEMCHEEMYGNEEGMVHEYVDEDDGYEDYYYDYAVEEEEEE
                     EEEDDIEATGEENKEEEDPKGIPDFWLTVLKNVDTLTPLIKKYDEPILKLLTDIKVKL
                     SDPGEPLSFTLEFHFKPNEYFKNELLTKTYVLKSKLAYYDPHPYRGTAIEYSTGCEID
                     WNEGKNVTLKTIKKKQKHRIWGTIRTVTEDFPKDSFFNFFSPHGITSNGRDGNDDFLL
                     GHNLRTYIIPRSVLFFSGDALESQQEGVVREVNDAIYDKIIYDNWMAAIEEVKACCKN
                     LEALVEDIDR"
BASE COUNT       742 a     432 c     589 g     752 t
ORIGIN

  Query Match             23.4%;  Score 356;  DB 9;  Length 2515;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-65;
  Matches  356;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1165 TGATAGCTGCGTAATCATACTGCGGCACCGTTTTTTTCTTGCAGCAGTAGCTGCTTGCGG 1224
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TGATAGCTGCGTAATCATACTGCGGCACCGTTTTTTTCTTGCAGCAGTAGCTGCTTGCGG 60

Qy   1225 AGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCGGCTCTCTCCTGCAGGATCGGTCAA 1284
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCGGCTCTCTCCTGCAGGATCGGTCAA 120

Qy   1285 CGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGTCGCCACTGCTTCAGTCCGGTTCTCA 1344
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGTCGCCACTGCTTCAGTCCGGTTCTCA 180

Qy   1345 AAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTGGATCGTCGTTCCCTCAGTCCGGAC 1404
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTGGATCGTCGTTCCCTCAGTCCGGAC 240

Qy   1405 GCCACTGCTAGGTCCGACCACCGCCGCTTCTGATATTTCGGTGAGTCTTTTCCTGTGGAG 1464
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCCACTGCTAGGTCCGACCACCGCCGCTTCTGATATTTCGGTGAGTCTTTTCCTGTGGAG 300

Qy   1465 GTTTGGTCTCCCGATCTCTGTGGTAGCCACCTTAGGCGTGTACGGTCCTTTGAAAA 1520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTTTGGTCTCCCGATCTCTGTGGTAGCCACCTTAGGCGTGTACGGTCCTTTGAAAA 356


RESULT 6
BC026325
LOCUS       BC026325                2524 bp    mRNA    linear   PRI 08-APR-2002
DEFINITION  Homo sapiens, nucleosome assembly protein 1-like 2, clone MGC:26243
            IMAGE:4798964, mRNA, complete cds.
ACCESSION   BC026325
VERSION     BC026325.1  GI:20070943
KEYWORDS    MGC.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2524)
  AUTHORS   Strausberg,R.
```

```
TITLE     Direct Submission
JOURNAL   Submitted (02-APR-2002) National Institutes of Health, Mammalian
          Gene Collection (MGC), Cancer Genomics Office, National Cancer
          Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
          USA
REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
          cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
          Toshiyuki and Piero Carninci (RIKEN)
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Sequencing Group at the Stanford Human Genome
          Center, Stanford University School of Medicine, Stanford, CA  94305
          Web site:       http://www-shgc.stanford.edu
          Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
          Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
          R. M.

          Clone distribution: MGC clone distribution information can be found
          through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
          Series: IRAK Plate: 32 Row: g Column: 23
          This clone was selected for full length sequencing because it
          passed the following selection criteria: matched mRNA gi: 11415047.
FEATURES             Location/Qualifiers
     source          1. .2524
                     /organism="Homo sapiens"
                     /db_xref="LocusID:4674"
                     /db_xref="taxon:9606"
                     /clone="MGC:26243 IMAGE:4798964"
                     /tissue_type="Brain, hypothalamus"
                     /clone_lib="NIH_MGC_96"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
     CDS             305. .1687
                     /codon_start=1
                     /product="nucleosome assembly protein 1-like 2"
                     /protein_id="AAH26325.1"
                     /db_xref="GI:20070944"
                     /translation="MAKSENRKELSESSQEEAGNQIMVEGLGEHLERGEDAAAGLGDD
                     GKCGEEAAAGLGEEGENGEDTAAGSGEDGKKGGDTDEDSEADRPKGLIGYVLDTDFVE
                     SLPVKVKYRVLALKKLQTRAANLESKFLREFHDIERKFAEMYQPLLEKRRQIINAIYE
                     PTEEECEYKSDSEDCDDEEMCHEEMYGNEEGMVHEYVDEDDGYEDYYYDYAVEEEEEE
                     EEEDDIEATGEENKEEEDPKGIPDFWLTVLKNVDTLTPLIKKYDEPILKLLTDIKVKL
                     SDPGEPLSFTLEPHFKPNEYFKNELLTKTYVLKSKLAYYDPHPYRGTAIEYSTGCEID
                     WNEGKNVTLKTIKKKQKHRIWGTIRTVTEDFPKDSFFNFFSPHGITSNGRDGNDDFLL
                     GHNLRTYIIPRSVLFFSGDALESQQEGVVREVNDAIYDKIIYDNWMAAIEEVKACCKN
                     LEALVEDIDR"
BASE COUNT      790 a    420 c    577 g    737 t
ORIGIN

  Query Match           19.7%;  Score 299.2;  DB 9;  Length 2524;
  Best Local Similarity  99.0%;  Pred. No. 3.8e-53;
  Matches  301;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy   1217 GCTTGCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCGGCTCTCTCCTGCAGGA 1276
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GCTTGCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCGGCTCTCTCCTGCAGGA 60

Qy   1277 TCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGTCGCCACTGCTTCAGTCC 1336
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGTCGCCACTGCTTCAGTCC 120

Qy   1337 GGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTGGATCGTCGTTCCCTCA 1396
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db    121 GGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTGGATCGTCGTTCACTCA 180

Qy   1397 GTCCGGACGCCACTGCTAGGTCCGACCACCGCCGCTTCTGATATTTCGGTGAGTCTTTTC 1456
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GTCCGGACGCCACTACTAGGTCCGACCACCGCCGCTTCTGATATTTCGGTGAGTCTTTTC 240

Qy   1457 CTGTGGAGGTTTGGTCTCCCGATCTCTGTGGTAGCCACCTTAGGCGTGTACGGTCCTTTG 1516
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||

Db    241 CTGTGGAGGTTTGGTCTCCCGATCCCTGTGGTAGCCACCTTAGGCGTGTACGGTCCTTTG 300

Qy   1517 AAAA 1520
          ||||
Db    301 AAAA 304


RESULT 7
HS84H6R
LOCUS       HS84H6R                  272 bp    DNA     linear   PRI 22-OCT-1995
DEFINITION  H.sapiens CpG island DNA genomic Mse1 fragment, clone 84h6, reverse
            read cpg84h6.rt1a.
ACCESSION   Z63470
VERSION     Z63470.1  GI:1035848
KEYWORDS    CpG island; genomic Mse1 fragment.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 272)
  AUTHORS   MacDonald,M., Huckle,E., Wilkinson,P. and Micklem,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1RQ, England. E-mail contact: humquery@sanger.ac.uk
REFERENCE   2  (bases 1 to 272)
  AUTHORS   Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.
  TITLE     Purification of CpG islands using a methylated DNA binding column
  JOURNAL   Nat. Genet. 6 (3), 236-244 (1994)
  MEDLINE   94282070
   PUBMED   8012384
COMMENT     Vector: pGEM-5Zf(-)
            Clones are available from the UK MRC Human Genome Mapping Project
            Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:
            http://www.hgmp.mrc.ac.uk/ for details
            or contact: biohelp@hgmp.mrc.ac.uk.
FEATURES             Location/Qualifiers
     source          1. .272
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="84h6"
                     /sex="male"
                     /tissue_type="blood"
                     /clone_lib="CGI-1"
                     /dev_stage="adult"
BASE COUNT       84 a     37 c     84 g     57 t     10 others
ORIGIN

  Query Match           17.2%;  Score 262;  DB 9;  Length 272;
  Best Local Similarity  96.3%;  Pred. No. 2.5e-45;
  Matches  262;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy    902 AAAGTTGGATGAATGAACTAATGAACGCAATGGGATTCCAGAAAACTCTGCGGGAGATGG 961
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db      1 AAAGTTGGATGAATGAACTAATGAACGCAATGGNATTCCAGAAAACTCTGCGGGAGATGG 60

Qy    962 GCTAGAGGACGAGGAGGAGGTGGATGAATCAGCCATGTTAGAGAGCCTGGGAAGGTGAGC 1021
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCTAGAGGACGAGGAGGAGGTGGATGAATCAGCCATGTTAGAGAGCCTGGGAAGGTGAGC 120

Qy   1022 AGAGTTGAAAACTTGATAGATCTAATAATTTACTGGCTCTGGGTTTGTCAGTCACTACAT 1081
          ||||| || |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGAGTNGANAACTTGATAGATCTAATAATTTACTGGCTCTGGGTNTGTCAGTCACTACAT 180

Qy   1082 TGCAGCAAATGAGATTAGAGCATAGTTGTGGGAGGGAAGGAGGTGACGCAGCAATCTATT 1141
          |||||||||||||||||| | ||||||||||||||||||||||||||||||||||||||
Db    181 TGCAGCAAATGAGATTAGAGCATAGTNGNGGGAGGGAAGGAGGTGACGCAGCAATCTATT 240

Qy   1142 TGCACCTAGAAATTTTAGGCAAGTGATAGCTG 1173
          |||||||||  | ||||||||||||||||||
Db    241 NGCACCTAGAANTNNTAGGCAAGTGATAGCTG 272


RESULT 8
```

```
AX327799
LOCUS       AX327799                1725 bp    DNA     linear   PAT 07-JAN-2002
DEFINITION  Sequence 1 from Patent WO0185995.
ACCESSION   AX327799
VERSION     AX327799.1  GI:18098072
KEYWORDS    .
SOURCE      Mus sp.
  ORGANISM  Mus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Avner,P., Rogner,U.C., Spyropoulos,D. and Rougeulle,C.
  TITLE     Identification of neural defects associated with the nucleosomal
            assembly protein 112 gene
  JOURNAL   Patent: WO 0185995-A 1 15-NOV-2001;
            INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE
            SCIENTIFIQUE (CNRS) (FR)
FEATURES             Location/Qualifiers
     source          1..1725
                     /organism="Mus sp."
                     /db_xref="taxon:10095"
BASE COUNT      428 a    443 c    391 g    463 t
ORIGIN

  Query Match             15.4%;  Score 233.8;  DB 6;  Length 1725;
  Best Local Similarity   65.1%;  Pred. No. 2.8e-39;
  Matches  486;  Conservative    0;  Mismatches  222;  Indels   39;  Gaps    8;

Qy    580 CTCCCCTCACAGGAAATGTTTCGAGAATTTTTCAACCTAAAATCATATAGCTTGTGAAAA 639
Db     74 CTTCCTCATCAGGAAATATTATGAGAATTTTCCCATTTAAAATCACACAGGTTGTGAAAA 133

Qy    640 ATACCGACAAACATAATATAGAATATTTAAATAACTGAC----ACGCCACCTAAAGACCA 695
Db    134 TTACAGA-AACCAGGGTACAGAATATTTAAACCACTGTCAGTTACATCATCCAAAGGCCA 192

Qy    696 TCAGTGCTAATTCCTGGTGTTTTTAATCTTTGAAGCGTTTGTTTATCAGCTCTTCCACCA 755
Db    193 CCTATGCTTATTTTTGGTAATTTTAAACCTCAAAGGATCTCTTTGTGGGCTCCTCCACTA 252

Qy    756 TCCACCTCTCCCCTCCCCAGGTCCCCGATCTAAAATCAAAGAGATTGATTTAGGATGGGT 815
Db    253 CCCTCCTCT--CTTTCCCAGAGCCTCAGGTTATAACCAAAGGGATAGACTAAAGACAATC 310

Qy    816 GGGTGCCTTGTCTTCTCTCATTGTTCGACATTTTAGTTACGTTTTCTCTGAGCTCTCTGG 875
Db    311 CAGTACCTTGCCCATTTTTTTCATTC-----CTTGTCACTGTTTCCATATAGCTCTTTTG 365

Qy    876 AAAGCATAAAAGTATAATATCTGTTAAAAGTTGGATGAATGAACTAATGAACGCAATGGG 935
Db    366 AAATTATGAACATATAGTATCAGTTGAAAACGGAATGAATGA-----------TACTGCA 414

Qy    936 ATTCCAGAAAACTCTGCGGGAGATGGGCTAGAGGACGAGGAGGAGGTGGATGAATCAGCC 995
Db    415 TTTCTGCAAAATTCCACAGGCTATAGGGTGGAAGATGAGCCATAGGTGGAGGAATCAGCC 474

Qy    996 ATGTTAGAGAGCCTGGGAAGGTGAGCAGAGTTGAAAACTTGATAGATCTAATAATTTACT 1055
Db    475 ATATTAGAGAATCTGGGAAGGCAAGAGGTGTTGAAATTTTGATTCATCTACTAATTTACT 534

Qy   1056 GGCTCTGGGTTTGTCAGTCACTACATTGCAGCAAATGAGATTAGAGCATAGTTGTGGGAG 1115
Db    535 GGCTCAGGATTTGTCAATCACTGCAGCCTGGCAAATGAGATTAGAGAAGAGTCCTGGGAG 594

Qy   1116 GGAAGGAGGTGACGCAGCAATCTATTTGCACCT-AGAAATTTTAGGCAAGTGATAGCTGC 1174
Db    595 GGAAGG-GGTGACGCAGCAACCTGCATACACTTAAAAAAAAAGAGCTGAGAGACAACTGC 653

Qy   1175 GTAATCATACTGCGGCACCGTTTTTTTCTT-------------GCAGCAGTAGCTGCTT 1220
Db    654 GTAATCATACTGCGGCACCAGTTCCTCCATCCCTCCGCCCCCGAGTGGCTGGAGCAGCTG 713

Qy   1221 GCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCGGCTCTCTCCTGCAGGATCGG 1280
Db    714 CTTGCGGAGGTCTGCCCACTGCGGCTCTCTGCAGTCTCTAGCCTGTTCCTTCAGGGCCTA 773

Qy   1281 TCAACGCAGCCGTCGCCGCCCTCTGCA 1307
Db    774 GAGTCTCCGCCCAGACAGCCGGTTTCA 800

RESULT 9
AL773526/c
LOCUS       AL773526              210614 bp    DNA     linear   HTG 17-AUG-2002
DEFINITION  Mus musculus chromosome X clone RP23-181M13, *** SEQUENCING IN
            PROGRESS ***, 5 unordered pieces.
ACCESSION   AL773526
VERSION     AL773526.5  GI:22204627
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 210614)
  AUTHORS   Heath,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 11, 2002 this sequence version replaced gi:21621733.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bM181M13
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 209617 bases at least Q40
            Consensus quality: 209945 bases at least Q30
            Consensus quality: 210088 bases at least Q20
            Insert size: 210214; sum-of-contigs
            Insert size: 194526; 6.8% error; agarose-fp
            Quality coverage: 9.25x in Q20 bases; sum-of-contigs Quality
            coverage: 10.05x in Q20 bases; agarose-fp
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *          1      23338: contig of 23338 bp in length
            *      23339  23438: gap of        100 bp
            *      23439      99523: contig of 76085 bp in length
            *      99524  99623: gap of        100 bp
            *      99624     144306: contig of 44683 bp in length
            *     144307 144406: gap of        100 bp
            *     144407     201734: contig of 57328 bp in length
            *     201735 201834: gap of        100 bp
            *     201835     210614: contig of 8780 bp in length.
FEATURES             Location/Qualifiers
     source          1..210614
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="X"
                     /clone="RP23-181M13"
                     /clone_lib="RPCI-23"
     misc_feature    1..23338
                     /note="assembly_fragment:01936
                     fragment_chain:1"
     misc_feature    23439..99523
```

```
                     /note="assembly_fragment:01297
                     fragment_chain:1"
     misc_feature    99624..144306
                     /note="assembly_fragment:02045
                     fragment_chain:2"
     misc_feature    144407..201734
                     /note="assembly_fragment:01852
                     fragment_chain:2"
     misc_feature    201835..210614
                     /note="assembly_fragment:03448"
BASE COUNT    61616 a  43211 c  44138 g  61249 t    400 others
ORIGIN

  Query Match             15.4%;  Score 233.8;  DB 2;  Length 210614;
  Best Local Similarity   65.1%;  Pred. No. 3.8e-39;
  Matches  486;  Conservative    0;  Mismatches  222;  Indels   39;  Gaps    8;

Qy      580 CTCCCCTCACAGGAAATGTTTCGAGAATTTTTCAACCTAAAATCATATAGCTTGTGAAAA 639
Db    52001 CTTCCTCATCAGGAAATATTATGAGAATTTTCCCATTTAAAATCACACAGGTTGTGAAAA 51942

Qy      640 ATACCGACAAACATAATATAGAATATTTAAATAACTGAC----ACGCCACCTAAAGACCA 695
Db    51941 TTACAGA-AACCAGGGTACAGAATATTTAAACCACTGTCAGTTACATCATCCAAAGGCCA 51883

Qy      696 TCAGTGCTAATTCCTGGTGTTTTTAATCTTTGAAGCGTTTGTTATCAGCTCTTCCACCA 755
Db    51882 CCTATGCTTATTTTTGGTAATTTTAAACCTCAAAGGATCTCTTTGTGGGCTCCTCCACTA 51823

Qy      756 TCCACCTCTCCCCTCCCCAGGTCCCCGATCTAAAATCAAAGAGATTGATTTAGGATGGGT 815
Db    51822 CCCTCCTCT--CTTTCCCAGAGCCTCAGGTTATAACCAAAGGGATAGACTAAAGACAATC 51765

Qy      816 GGGTGCCTTGTCTTCTCTCATTGTTCGACATTTTAGTTACGTTTTCTCTGAGCTCTCTGG 875
Db    51764 CAGTACCTTGCCCATTTTTTTCATTC-----CTTGTCACTGTTTCCATATAGCTCTTTTG 51710

Qy      876 AAAGCATAAAAGTATAATATCTGTTAAAAGTTGGATGAATGAACTAATGAACGCAATGGG 935
Db    51709 AAATTATGAACATATAGTATCAGTTGAAAACGGAATGAATGA-----------TACTGCA 51661

Qy      936 ATTCCAGAAAACTCTGCGGGAGATGGGCTAGAGGACGAGGAGGAGGTGGATGAATCAGCC 995
Db    51660 TTTCTGCAAAATTCCACAGGCTATAGGGTGGAAGATGAGCCATAGGTGGAGGAATCAGCC 51601

Qy      996 ATGTTAGAGAGCCTGGGAAGGTGAGCAGAGTTGAAAACTTGATAGATCTAATAATTTACT 1055
Db    51600 ATATTAGAGAATCTGGGAAGGCAAGAGGTGTTGAAATTTTGATTCATCTACTAATTTACT 51541

Qy     1056 GGCTCTGGGTTTGTCAGTCACTACATTGCAGCAAATGAGATTAGAGCATAGTTGTGGGAG 1115
Db    51540 GGCTCAGGATTTGTCAATCACTGCAGCCTGGCAAATGAGATTAGAGAAGAGTCCTGGGAG 51481

Qy     1116 GGAAGGAGGTGACGCAGCAATCTATTTGCACCT-AGAAATTTTAGGCAAGTGATAGCTGC 1174
Db    51480 GGAAGG-GGTGACGCAGCAACCTGCATACACTTAAAAAAAAAGAGCTGAGAGACAACTGC 51422

Qy     1175 GTAATCATACTGCGGCACCGTTTTTTTCTT--------------GCAGCAGTAGCTGCTT 1220
Db    51421 GTAATCATACTGCGGCACCAGTTCCTCCATCCCTCCGCCCCCGAGTGGCTGGAGCAGCTG 51362

Qy     1221 GCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCGGCTCTCTCCTGCAGGATCGG 1280
Db    51361 CTTGCGGAGGTCTGCCCACTGCGGCTCTCTGCAGTCTCTAGCCTGTTCCTTCAGGGCCTA 51302

Qy     1281 TCAACGCAGCCGTCGCCGCCCTCTGCA 1307
Db    51301 GAGTCTCCGCCCAGACAGCCGGTTTCA 51275

RESULT 10
AJ421480
LOCUS       AJ421480              214384 bp    DNA     linear   ROD 12-JUN-2002
DEFINITION  Mus musculus X-inactivation center region; segment 3/3.
ACCESSION   AJ421480
VERSION     AJ421480.1  GI:21425587
KEYWORDS    Cdx4 gene; Napll2 gene; Ppnx gene.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Chureau,C., Prissette,M., Bourdet,A., Barbe,V., Cattolico,L.,
            Jones,L., Eggen,A., Avner,P. and Duret,L.
  TITLE     Comparative sequence analysis of the X-inactivation center region
            in mouse, human, and bovine
  JOURNAL   Genome Res. 12 (6), 894-908 (2002)
  MEDLINE   22040331
   PUBMED   12045143
REFERENCE   2  (bases 1 to 214384)
  AUTHORS   Duret,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-NOV-2001) Duret L., Biologie et Biologie Evolutive,
            CNRS - Universite Lyon 1, 16, rue Raphael Dubois, 69622
            Villeurbanne cedex, FRANCE
COMMENT     Sequenced by the Genoscope. Contact: Philip Avner, Institut
            Pasteur, 25, rue du  Docteur Roux  75015 Paris  France.
FEATURES             Location/Qualifiers
     source          1..214384
                     /organism="Mus musculus"
                     /strain="129/Sv"
                     /db_xref="taxon:10090"
                     /chromosome="X"
     repeat_region   complement(54..169)
                     /rpt_family="B4A SINE/B4"
     repeat_region   181..221
                     /note="T-rich Low_complexity"
     repeat_region   complement(222..264)
                     /rpt_family="B1_MM SINE/Alu"
     repeat_region   complement(473..643)
                     /rpt_family="Lx6 LINE/L1"
     repeat_region   647..888
                     /rpt_family="L1_MM LINE/L1"
     repeat_region   complement(936..1081)
                     /rpt_family="B1_MM SINE/Alu"
     repeat_region   complement(1196..1399)
                     /rpt_family="B2_Mm1 SINE/B2"
     repeat_region   complement(1449..1588)
                     /rpt_family="B1_MM SINE/Alu"
     repeat_region   1786..1855
                     /note="GA-rich Low_complexity"
     repeat_region   complement(2115..2262)
                     /rpt_family="B1_MM SINE/Alu"
     repeat_region   complement(2317..2439)
                     /rpt_family="L1M4 LINE/L1"
     repeat_region   2517..2581
                     /note="(TCTCTG)n Simple_repeat"
     repeat_region   complement(2584..2650)
                     /rpt_family="ID1_MM SINE/ID"
     repeat_region   3172..3220
                     /note="GA-rich Low_complexity"
     repeat_region   3346..3371
                     /note="(T)n Simple_repeat"
     repeat_region   complement(3372..3519)
                     /rpt_family="B1_MM SINE/Alu"
     repeat_region   complement(3901..4008)
                     /rpt_family="B1F SINE/Alu"
     repeat_region   complement(4009..4084)
                     /rpt_family="B4A SINE/B4"
     repeat_region   5092..5178
                     /note="(GA)n Simple_repeat"
     repeat_region   5304..5352
                     /rpt_family="MIR SINE/MIR"
     gene            complement(5774..14602)
                     /gene="Cdx4"
     mRNA            complement(join(5774..6564,7844..7989,13942..14602))
```

```
                     /gene="Cdx4"
CDS                  /note="match: cDNA L08061"
                     complement(join(6358. .6564,7844. .7989,13942. .14437))
                     /gene="Cdx4"
                     /codon_start=1
                     /protein_id="CAD33953.1"
                     /db_xref="GI:21425588"
                     /translation="MYGSCLLEKEAGMYPGTLRSPGGSSTAGVGTSGGSGSPLPASNF
                     TAAPVYPHYVGYPHMSNMDPHGPSLGAWSSPYSPPREDWSTYPGPPSTMGTVPMNDMT
                     SPVFGSPDYSTLGPTSGASNGGSLPDAASESLVSLDSGTSGATSPSRSRHSPYAWMRK
                     TVQVTGKTRTKEKYRVVYTDHQRLELEKEFHCNRYITIRRKSELAVNLGLSERQVKIW
                     FQNRRAKERKMIKKKISQFENTGGSVQSDSGSISPGELPNAFFTTPSAVRGFQPIEIQ
                     QVIVSE"
repeat_region        complement(6646. .6736)
                     /rpt_family="MIR SINE/MIR"
repeat_region        6963. .7114
                     /rpt_family="B1_MM SINE/Alu"
repeat_region        complement(8349. .8497)
                     /rpt_family="B1_MM SINE/Alu"
repeat_region        10659. .10704
                     /note="(CAAAA)n Simple_repeat"
repeat_region        11216. .11523
                     /rpt_family="L1_MM LINE/L1"
repeat_region        11542. .11601
                     /note="(CA)n Simple_repeat"
repeat_region        11724. .11743
                     /note="(TC)n Simple_repeat"
repeat_region        13704. .13727
                     /note="(CAAA)n Simple_repeat"
repeat_region        complement(15069. .15142)
                     /rpt_family="L2 LINE/L2"
repeat_region        15225. .15299
                     /note="(GAA)n Simple_repeat"
repeat_region        15533. .15722
                     /rpt_family="B2_Mm1 SINE/B2"
repeat_region        complement(15840. .15977)
                     /rpt_family="B1_MM SINE/Alu"
CDS                  complement(16474. .>16944)
                     /note="ribosomal protein L4 pseudogene
                     match: cDNA X82180"
                     /pseudo
                     /codon_start=1
                     /evidence=not_experimental
repeat_region        complement(17077. .17280)
                     /rpt_family="B3A SINE/B2"
CDS                  complement(<17385. .17858)
                     /note="ribosomal protein L4 pseudogene
                     match: cDNA X82180"
                     /pseudo
                     /codon_start=1
                     /evidence=not_experimental
repeat_region        complement(18450. .18521)
                     /rpt_family="RSINE1 SINE/B4"
repeat_region        18537. .18712
                     /rpt_family="B2_Mm1 SINE/B2"
repeat_region        complement(18719. .18806)
                     /rpt_family="RSINE1 SINE/B4"
repeat_region        18886. .18992
                     /note="(GAA)n Simple_repeat"
repeat_region        complement(19188. .19218)
                     /rpt_family="L2 LINE/L2"
repeat_region        19219. .19617
                     /rpt_family="MTA LTR/MaLR"
repeat_region        complement(19618. .19734)
                     /rpt_family="L2 LINE/L2"
repeat_region        19766. .19888
                     /rpt_family="B1F SINE/Alu"
repeat_region        complement(20092. .20195)
                     /rpt_family="MIR SINE/MIR"
repeat_region        20306. .20425
                     /rpt_family="B1F SINE/Alu"
repeat_region        20600. .20994
                     /rpt_family="MTA LTR/MaLR"
repeat_region        21055. .21158
                     /rpt_family="B1_MM SINE/Alu"
repeat_region        complement(21430. .21508)
                     /rpt_family="B1_MM SINE/Alu"
repeat_region        21765. .21802
                     /note="C-rich Low_complexity"
repeat_region        22116. .22153
                     /note="(TTTC)n Simple_repeat"
repeat_region        complement(22154. .22294)
                     /rpt_family="B1_MM SINE/Alu"
repeat_region        complement(22570. .22706)
                     /rpt_family="B1_MM SINE/Alu"
repeat_region        22719. .22764
                     /note="(T)n Simple_repeat"
repeat_region        complement(22765. .22912)
                     /rpt_family="B1_MM SINE/Alu"
repeat_region        complement(23235. .23359)
                     /rpt_family="ORR1B LTR/MaLR"
repeat_region        23390. .23958
                     /rpt_family="Lx6 LINE/L1"
repeat_region        complement(24392. .24517)
                     /rpt_family="Lx2 LINE/L1"
repeat_region        complement(24518. .24645)
                     /rpt_family="L1 LINE/L1"
repeat_region        25276. .25341
                     /note="(GA)n Simple_repeat"
repeat_region        25346. .25428
                     /note="(GGAA)n Simple_repeat"
repeat_region        complement(25615. .25932)
                     /rpt_family="L1 LINE/L1"
repeat_region        25932. .26590
                     /rpt_family="L1_MM LINE/L1"
repeat_region        complement(26830. .27016)
                     /rpt_family="ORR1D LTR/MaLR"
repeat_region        27142. .27217
                     /rpt_family="MTD LTR/MaLR"
repeat_region        27218. .27364
                     /rpt_family="B1_MM SINE/Alu"
repeat_region        27408. .27536
                     /rpt_family="MTD LTR/MaLR"
repeat_region        complement(27537. .27576)
                     /rpt_family="ORR1A3 LTR/MaLR"
repeat_region        27575. .27666
                     /rpt_family="B1_MM SINE/Alu"
repeat_region        27675. .27826
                     /rpt_family="MTA LTR/MaLR"
repeat_region        27827. .28219
                     /rpt_family="MTA LTR/MaLR"
repeat_region        complement(28478. .28625)
                     /rpt_family="B1_MM SINE/Alu"
STS                  28674. .28823
                     /standard_name="DXPas28"
repeat_region        28721. .28794
```

Query Match 15.4%; Score 233.8; DB 10; Length 214384;
Best Local Similarity 65.1%; Pred. No. 3.8e-39;
Matches 486; Conservative 0; Mismatches 222; Indels 39; Gaps 8;

```
Qy     580 CTCCCCTCACAGGAAATGTTTCGAGAATTTTTCAACCTAAAATCATATAGCTTGTGAAAA 639
           || ||          |||||||| ||  |||||||||| | |   |||||||| | || |||||||||
Db  161158 CTTCCTCATCAGGAAATATTATGAGAATTTTCCCATTTAAAATCACACAGGTTGTGAAAA 161217

Qy     640 ATACCGACAAACATAATATAGAATATTTAAATAACTGAC----ACGCCACCTAAAGACCA 695
           ||| || || ||   || ||||||||||||  |||| |    ||  || | |||| |||
Db  161218 TTACAGA-AACCAGGGTACAGAATATTTAAACCACTGTCAGTTACATCATCCAAAGGCCA 161276

Qy     696 TCAGTGCTAATTCCTGGTGTTTTTAATCTTTGAAGCGTTTGTTTATCAGCTCTTCCACCA 755
           | |||| |||  ||||  |||||| | |  |||   | | |||| |  |||| |||||| |
Db  161277 CCTATGCTTATTTTTGGTAATTTTAAACCTCAAAGGATCTCTTTGTGGGCTCCTCCACTA 161336

Qy     756 TCCACCTCTCCCCTCCCCAGGTCCCCGATCTAAAATCAAAGAGATTGATTTAGGATGGGT 815
           || |||||  | | ||||||  || |    || || ||||| ||| || | | ||
Db  161337 CCCTCCTCT--CTTTCCCAGAGCCTCAGGTTATAACCAAAGGGATAGACTAAAGACAATC 161394
```

```
Qy       816 GGGTGCCTTGTCTTCTCTCATTGTTCGACATTTTAGTTACGTTTTCTCTGAGCTCTCTGG 875
Db    161395 CAGTACCTTGCCCATTTTTTTCATTC-----CTTGTCACTGTTTCCATATAGCTCTTTTG 161449

Qy       876 AAAGCATAAAAGTATAATATCTGTTAAAAGTTGGATGAATGAACTAATGAACGCAATGGG 935
Db    161450 AAATTATGAACATATAGTATCAGTTGAAAACGGAATGAATGA-----------TACTGCA 161498

Qy       936 ATTCCAGAAAACTCTGCGGGAGATGGGCTAGAGGACGAGGAGGAGGTGGATGAATCAGCC 995
Db    161499 TTTCTGCAAAATTCCACAGGCTATAGGGTGGAAGATGAGCCATAGGTGGAGGAATCAGCC 161558

Qy       996 ATGTTAGAGAGCCTGGGAAGGTGAGCAGAGTTGAAAACTTGATAGATCTAATAATTTACT 1055
Db    161559 ATATTAGAGAATCTGGGAAGGCAAGAGGTGTTGAAATTTTGATTCATCTACTAATTTACT 161618

Qy      1056 GGCTCTGGGTTTGTCAGTCACTACATTGCAGCAAATGAGATTAGAGCATAGTTGTGGGAG 1115
Db    161619 GGCTCAGGATTTGTCAATCACTGCAGCCTGGCAAATGAGATTAGAGAAGAGTCCTGGGAG 161678

Qy      1116 GGAAGGAGGTGACGCAGCAATCTATTTGCACCT-AGAAATTTTAGGCAAGTGATAGCTGC 1174
Db    161679 GGAAGG-GGTGACGCAGCAACCTGCATACACTTAAAAAAAAAGAGCTGAGAGACAACTGC 161737

Qy      1175 GTAATCATACTGCGGCACCGTTTTTTTCTT--------------GCAGCAGTAGCTGCTT 1220
Db    161738 GTAATCATACTGCGGCACCAGTTCCTCCATCCCTCCGCCCCCGAGTGGCTGGAGCAGCTG 161797

Qy      1221 GCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCGGCTCTCTCCTGCAGGATCGG 1280
Db    161798 CTTGCGGAGGTCTGCCCACTGCGGCTCTCTGCAGTCTCTAGCCTGTTCCTTCAGGGCCTA 161857

Qy      1281 TCAACGCAGCCGTCGCCGCCCTCTGCA 1307
Db    161858 GAGTCTCCGCCCAGACAGCCGGTTTCA 161884
```

```
RESULT 11
AX327800
LOCUS       AX327800                2819 bp    DNA     linear   PAT 07-JAN-2002
DEFINITION  Sequence 2 from Patent WO0185995.
ACCESSION   AX327800
VERSION     AX327800.1  GI:18098073
KEYWORDS    .
SOURCE      Mus sp.
  ORGANISM  Mus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Avner,P., Rogner,U.C., Spyropoulos,D. and Rougeulle,C.
  TITLE     Identification of neural defects associated with the nucleosomal
            assembly protein 112 gene
  JOURNAL   Patent: WO 0185995-A 2 15-NOV-2001;
            INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE
            SCIENTIFIQUE (CNRS) (FR)
FEATURES             Location/Qualifiers
     source          1..2819
                     /organism="Mus sp."
                     /db_xref="taxon:10095"
BASE COUNT     827 a    522 c    670 g    800 t
ORIGIN

  Query Match             11.0%;  Score 166.8;  DB 6;  Length 2819;
  Best Local Similarity   64.9%;  Pred. No. 4.6e-25;
  Matches  322;  Conservative    0;  Mismatches  147;  Indels   27;  Gaps    4;

Qy       827 CTTCTCTCATTGTTCGACATTTTAGTTACGTTTTCTCTGAGCTCTCTGGAAAGCATAAAA 886
Db         5 CTTGCCCATTTTTTTCATTCCTTGTCACTGTTTCCATATAGCTCTTTTGAAATTATGAAC 64

Qy       887 GTATAATATCTGTTAAAAGTTGGATGAATGAACTAATGAACGCAATGGGATTCCAGAAAA 946
Db        65 ATATAGTATCAGTTGAAAACGGAATGAATGA-----------TACTGCATTTCTGCAAAA 113

Qy       947 CTCTGCGGGAGATGGGCTAGAGGACGAGGAGGAGGTGGATGAATCAGCCATGTTAGAGAG 1006
Db       114 TTCCACAGGCTATAGGGTGGAAGATGAGCCATAGGTGGAGGAATCAGCCATATTAGAGAA 173

Qy      1007 CCTGGGAAGGTGAGCAGAGTTGAAAACTTGATAGATCTAATAATTTACTGGCTCTGGGTT 1066
Db       174 TCTGGGAAGGCAAGAGGTGTTGAAATTTTGATTCATCTACTAATTTACTGGCTCAGGATT 233

Qy      1067 TGTCAGTCACTACATTGCAGCAAATGAGATTAGAGCATAGTTGTGGGAGGGAAGGAGGTG 1126
Db       234 TGTCAATCACTGCAGCCTGGCAAATGAGATTAGAGAAGAGTCCTGGGAGGGAAGG-GGTG 292

Qy      1127 ACGCAGCAATCTATTTGCACCT-AGAAATTTTAGGCAAGTGATAGCTGCGTAATCATACT 1185
Db       293 ACGCAGCAACCTGCATACACTTAAAAAAAAAGAGCTGAGAGACAACTGCGTAATCATACT 352

Qy      1186 GCGGCACCGTTTTTTTCTT--------------GCAGCAGTAGCTGCTTGCGGAGGAGGT 1231
Db       353 GCGGCACCAGTTCCTCCATCCCTCCGCCCCCGAGTGGCTGGAGCAGCTGCTTGCGGAGGT 412

Qy      1232 CTGCCCACTGCAGCTCTCTGCAGTCTCCGGCTCTCTCCTGCAGGATCGGTCAACGCAGCC 1291
Db       413 CTGCCCACTGCGGCTCTCTGCAGTCTCTAGCCTGTTCCTTCAGGGCCTAGAGTCTCCGCC 472

Qy      1292 GTCGCCGCCCTCTGCA 1307
Db       473 CAGACAGCCGGTTTCA 488
```

```
RESULT 12
MIDVTRN/c
LOCUS       MIDVTRN                 2226 bp    DNA     linear   INV 30-NOV-1997
DEFINITION  D. virilis mitochondrial DNA for origin of replication, small
            mitochondrial ribosomal RNA, transfer tRNAs tRNA-fMet, tRNA-Gln and
            tRNA-Ile.
ACCESSION   X05914
VERSION     X05914.1  GI:12916
KEYWORDS    AT rich region; mitochondrial ribosomal RNA; NADH dehydrogenase;
            origin of replication; small ribosomal RNA; transfer RNA; transfer
            RNA-Gln; transfer RNA-Ile; transfer RNA-Met; transfer RNA-Val.
SOURCE      Drosophila virilis.
  ORGANISM  Mitochondrion Drosophila virilis
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 2226)
  AUTHORS   Clary,D.O. and Wolstenholme,D.R.
  TITLE     Drosophila mitochondrial DNA: conserved sequences in the A + T-rich
            region and supporting evidence for a secondary structure model of
            the small ribosomal RNA
  JOURNAL   J. Mol. Evol. 25 (2), 116-125 (1987)
  MEDLINE   88011348
   PUBMED   3116271
COMMENT     mitochondrial gene: tga codes for Trp !.
FEATURES             Location/Qualifiers
     source          1..2226
                     /organism="Drosophila virilis"
                     /organelle="mitochondrion"
                     /strain="stock 2375.8 (Chile)"
                     /db_xref="taxon:7244"
                     /clone_lib="pBR322"
     CDS             complement(<1..169)
                     /note="ND2 gene product (AA 1 - 56) (1 is 1st base in
                     codon)"
                     /codon_start=1
                     /transl_table=5
                     /protein_id="CAA29341.1"
                     /db_xref="GI:578700"
                     /db_xref="SPTREMBL:Q34398"
                     /translation="MFNNSSKILFITMMIIGTLITVTSNSWLGAWMGLEINLLSFIPL
                     MSDNNNFKSTEA"
```

```
RESULT 2
AA324132
LOCUS       AA324132                 367 bp    mRNA    linear   EST 20-APR-1997
DEFINITION  EST27025 Cerebellum II Homo sapiens cDNA 5' end, mRNA sequence.
ACCESSION   AA324132
VERSION     AA324132.1  GI:1976479
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 367)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
            ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He
            , Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,
            Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymon
            Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., 
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patt
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human (
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1..367
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):124708"
                     /db_xref="taxon:9606"
                     /clone_lib="Cerebellum II"
                     /tissue_type="cerebellum"
                     /dev_stage="adult"
                     /note="Organ: brain; Vector: pBluescript SK-; Site_1
                     EcoRI; Site_2: XhoI"
BASE COUNT       54 a    113 c     92 g     98 t     10 others
ORIGIN

  Query Match             22.7%;  Score 345.4;  DB 9;  Length 367;
  Best Local Similarity   96.9%;  Pred. No. 8.9e-68;
  Matches  346;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps

Qy       1164 GTGATAGCTGCGTAATCATACTGCGGCACCGTTTTTTTCTTGCAGCAGTAGCTGCTTGCG 12:
              |||||||||||||||||||||||||||||||||||||| ||||||||||||||| |||||
Db          1 GTGATAGCTGCGTAATCATACTGCGGCACCGTTTTTTTNTTGCAGCAGTAGCTGTTTGCG 60

Qy       1224 GAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCGGCTCTCTCCTGCAGGATCGGTCA 12(
              |||||||||||||||||||||||||||||||||||||||||| | ||||||| |||||||
Db         61 GAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCGGCTCTNTNCTGCAGGNTCGGTCA 12(

Qy       1284 ACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGTCGCCACTGCTTCAGTCCGGTTCTC 13'
              |||||||||||||||||||||||||||||||||||||||||||| |||| ||||||||||
Db        121 ACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGTCGCCACTGNTTCANTCCGGTTCTC 18(

Qy       1344 AAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTGGATCGTCGTTCCCTCAGTCCGGA 14(
              |||||||||||||||||||||||||||||||||||||| |||| ||||||||||||||||
Db        181 AAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTGGNTCGTNGTTCCCTCAGTCCGGA 24(

Qy       1404 CGCCACTGCTAGGTCCGACCACCGCCGCTTCTGATATTTCGGTGAGTCTTTTCCTGTGGA 14(
              |||||||||||||||||||||||| ||||||| |||||||||||||||||||||||||||
Db        241 CGCCACTGCTAGGTCCGACCACCGNCGCTTCTNATATTTCGGTGAGTCTTTTCCTGTGGA 30(

Qy       1464 GGTTTGGTCTCCCGATCTCTGTGGTAGCCACCTTAGGCGTGTACGGTCCTTTGAAAA 1520
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGTTTGGTCTCCCGATCTCTGTGGTAGCCACCTTAGGCGTGTACGGTCCTTTGAAAA 357

RESULT 3
AL120587
```

```
Db    241 CTGTGGAGGTTTGGTCTCCCGATCCCTGTGGTAGCCACCTTAGGCGTGTACGGTCCTTTG 300

Qy   1517 AAAA 1520
          ||||
Db    301 AAAA 304


RESULT 7
HS84H6R
LOCUS       HS84H6R                  272 bp    DNA     linear   PRI 22-OCT-1995
DEFINITION  H.sapiens CpG island DNA genomic MseI fragment, clone 84h6, reverse
            read cpg84h6.rt1a.
ACCESSION   Z63470
VERSION     Z63470.1  GI:1035848
KEYWORDS    CpG island; genomic MseI fragment.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 272)
  AUTHORS   MacDonald,M., Huckle,E., Wilkinson,P. and Micklem,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1RQ, England. E-mail contact: humquery@sanger.ac.uk
REFERENCE   2  (bases 1 to 272)
  AUTHORS   Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.
  TITLE     Purification of CpG islands using a methylated DNA binding column
  JOURNAL   Nat. Genet. 6 (3), 236-244 (1994)
  MEDLINE   94282070
    PUBMED  8012384
COMMENT     Vector: pGEM-5Zf(-)
            Clones are available from the UK MRC Human Genome Mapping Project
            Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:
            http://www.hgmp.mrc.ac.uk/ for details
            or contact: biohelp@hgmp.mrc.ac.uk.
FEATURES             Location/Qualifiers
     source          1. .272
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="84h6"
                     /sex="male"
                     /tissue_type="blood"
                     /clone_lib="CGI-1"
                     /dev_stage="adult"
BASE COUNT       84 a     37 c     84 g     57 t     10 others
ORIGIN

  Query Match          17.2%;  Score 262;  DB 9;  Length 272;
  Best Local Similarity   96.3%;  Pred. No. 2.5e-45;
  Matches  262;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy    902 AAAGTTGGATGAATGAACTAATGAACGCAATGGGATTCCAGAAAACTCTGCGGGAGATGG 961
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db      1 AAAGTTGGATGAATGAACTAATGAACGCAATGGNATTCCAGAAAACTCTGCGGGAGATGG 60

Qy    962 GCTAGAGGACGAGGAGGAGGTGGATGAATCAGCCATGTTAGAGAGCCTGGGAAGGTGAGC 1021
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCTAGAGGACGAGGAGGAGGTGGATGAATCAGCCATGTTAGAGAGCCTGGGAAGGTGAGC 120

Qy   1022 AGAGTTGAAAACTTGATAGATCTAATAATTTACTGGCTCTGGGTTTGTCAGTCACTACAT 1081
          ||||| || ||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    121 AGAGTNGANAACTTGATAGATCTAATAATTTACTGGCTCTGGGTNTGTCAGTCACTACAT 180

Qy   1082 TGCAGCAAATGAGATTAGAGCATAGTTGTGGGAGGGAAGGAGGTGACGCAGCAATCTATT 1141
          |||||||||||||||||||||| | ||||||||||||||||||||||||||||||||||
Db    181 TGCAGCAAATGAGATTAGAGCATAGTNGNGGGAGGGAAGGAGGTGACGCAGCAATCTATT 240

Qy   1142 TGCACCTAGAAATTTTAGGCAAGTGATAGCTG 1173
          ||||||||| | ||||||||||||||||
Db    241 NGCACCTAGAANTNNTAGGCAAGTGATAGCTG 272


RESULT 8
```

```
     1081 TTGCAGCAAATGAGATTAGAGCATAGTTGTGGGAGGGAAGGAGGTGACGCAGCAATCTAT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     1081 TTGCAGCAAATGAGATTAGAGCATAGTTGTGGGAGGGAAGGAGGTGACGCAGCAATCTAT 1140

     1141 TTGCACCTAGAAATTTTAGGCAAGTGATAGCTGCGTAATCATACTGCGGCACCGTTTTTT 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     1141 TTGCACCTAGAAATTTTAGGCAAGTGATAGCTGCGTAATCATACTGCGGCACCGTTTTTT 1200

     1201 TCTTGCAGCAGTAGCTGCTTGCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     1201 TCTTGCAGCAGTAGCTGCTTGCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCG 1260

     1261 GCTCTCTCCTGCAGGATCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGTC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     1261 GCTCTCTCCTGCAGGATCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGTC 1320

     1321 GCCACTGCTTCAGTCCGGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTG 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 GCCACTGCTTCAGTCCGGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTG 1380

Qy   1381 GATCGTCGTTCCCTCAGTCCGGACGCCACTGCTAGGTCCGACCACCGCCGCTTCTGATAT 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 GATCGTCGTTCCCTCAGTCCGGACGCCACTGCTAGGTCCGACCACCGCCGCTTCTGATAT 1440

Qy   1441 TTCGGTGAGTCTTTTCCTGTGGAGGTTTGGTCTCCCGATCTCTGTGGTAGCCACCTTAGG 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 TTCGGTGAGTCTTTTCCTGTGGAGGTTTGGTCTCCCGATCTCTGTGGTAGCCACCTTAGG 1500

Qy   1501 CGTGTACGGTCCTTTGAAAA 1520
          ||||||||||||||||||||
Db   1501 CGTGTACGGTCCTTTGAAAA 1520
```

RESULT 3
AC004074

LOCUS	AC004074 153578 bp DNA linear PRI 12-APR-1998
DEFINITION	Homo Sapiens Chromosome X clone bWXD759, complete sequence.
ACCESSION	AC004074
VERSION	AC004074.1 GI:3046270
KEYWORDS	HTG.
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 153578)
AUTHORS	Chen,E., Brownstein,B.H., States,D.J., Schlessinger,D. and Mazzarella,R:
TITLE	Direct Submission
JOURNAL	Unpublished (1997)
REFERENCE	2 (bases 1 to 153578)
AUTHORS	Brownstein,B.H., States,D.J. and Mazzarella,R.
TITLE	Direct Submission
JOURNAL	Submitted (29-JAN-1998) Center for Genetics in Medicine, Box 8232, Washington University School of Medicine, 4566 Scott Avenue, St. Louis, MO 63110, USA
REFERENCE	3 (bases 1 to 153578)
AUTHORS	Brownstein,B.H., States,D.J. and Mazzarella,R.
TITLE	Direct Submission
JOURNAL	Submitted (12-APR-1998) Center for Genetics in Medicine, Box 8232, Washington University School of Medicine, 4566 Scott Avenue, St. Louis, MO 63110, USA
COMMENT	On Apr 13, 1998 this sequence version replaced gi:2822135. Current status of this project is available at: 'http://www.ibc.wustl.edu/cgm/seq_projects.html' Submitted by: Ellson Chen, Advanced Center for Genetic Technology, Applied Biosystems Division of Perlin Elmer Corp., 850 Lincoln Center Drive, Foster City, CA 94404 USA e-mail: ellson@genseq.apldbio.com

and

Buddy Brownstein,
Center for Genetics in Medicine,
Washington University School of Medicine, Box 8232
4566 Scott Avenue,
St. Louis, MO 63110, USA
e-mail: buddy@genetics.wustl.edu

and

David J. States,
Institute for Biomedical Computing
Washington University in St. Louis
700 South Euclid Ave.
St. Louis, MO 63108 USA
e-mail: states@ibc.wustl.edu.

FEATURES Location/Qualifiers
source 1. .153578
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="bWXD759"

BASE COUNT 45069 a 29281 c 29096 g 50132 t
ORIGIN

Query Match 100.0%; Score 1520; DB 9; Length 153578;
Best Local Similarity 100.0%; Pred. No. 5.4e-312;
Matches 1520; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ACTTAAAGGAAAAATTTATCTATAAACTGACAGAATTTAGAAATAAATACAACAATATGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136351 ACTTAAAGGAAAAATTTATCTATAAACTGACAGAATTTAGAAATAAATACAACAATATGT 136410

Qy      61 AAACAGTTTTAATATCTGTGATAGTAACAAATTCTTTAAATCTGGAAAATAATAGTCACT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136411 AAACAGTTTTAATATCTGTGATAGTAACAAATTCTTTAAATCTGGAAAATAATAGTCACT 136470

Qy     121 TAAAATTTTAAAAAAATTGTTCAATTAATAAATGATCCAAGTTAGAAATATGAACAAAATA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136471 TAAAATTTTAAAAAAATTGTTCAATTAATAAATGATCCAAGTTAGAAATATGAACAAAATA 136530

Qy     181 AACCTCACCAATAATTACTATAGAGAGGAAATTTTAATTACTGCAAAGCTTTCCATCCTA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136531 AACCTCACCAATAATTACTATAGAGAGGAAATTTTAATTACTGCAAAGCTTTCCATCCTA 136590

Qy     241 TAAATACATTATCAAATAGTTTAACCATTTCTTTAATGCTGAGATTTAGATTATTTCCAA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136591 TAAATACATTATCAAATAGTTTAACCATTTCTTTAATGCTGAGATTTAGATTATTTCCAA 136650

Qy     301 TTAACTCAAAAGCATCAAGCAAATGTTATGATTTCTAAGAATAAACATAACTTTCCATTT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136651 TTAACTCAAAAGCATCAAGCAAATGTTATGATTTCTAAGAATAAACATAACTTTCCATTT 136710

Qy     361 TGGCTTTTGTATATATGTATATTTCTAACGGCTGTTAAAGCCAGCATTAAGAAGGAGAAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136711 TGGCTTTTGTATATATGTATATTTCTAACGGCTGTTAAAGCCAGCATTAAGAAGGAGAAG 136770

Qy     421 CAGAAAGTCAGTATTGGGACTGGGGTTATTTATAAGCCAGGCAACTGGTTAATTGTGGTT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136771 CAGAAAGTCAGTATTGGGACTGGGGTTATTTATAAGCCAGGCAACTGGTTAATTGTGGTT 136830

Qy     481 AATTGTCTGGTATGTTTACTAGTCACGTAGTTGTATACACCATACTAGTTTTTCATCACA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136831 AATTGTCTGGTATGTTTACTAGTCACGTAGTTGTATACACCATACTAGTTTTTCATCACA 136890

Qy     541 GGCCCTCATTCGCCCCCACTGCCATCGGACTTCCTCCTCCTCCCCTCACAGGAAATGTTT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136891 GGCCCTCATTCGCCCCCACTGCCATCGGACTTCCTCCTCCTCCCCTCACAGGAAATGTTT 136950

Qy     601 CGAGAATTTTTCAACCTAAAATCATATAGCTTGTGAAAAATACCGACAAACATAATATAG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136951 CGAGAATTTTTCAACCTAAAATCATATAGCTTGTGAAAAATACCGACAAACATAATATAG 137010

Qy     661 AATATTTAAATAACTGACACGCCACCTAAAGACCATCAGTGCTAATTCCTGGTGTTTTTA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137011 AATATTTAAATAACTGACACGCCACCTAAAGACCATCAGTGCTAATTCCTGGTGTTTTTA 137070

Qy     721 ATCTTTGAAGCGTTTGTTTATCAGCTCTTCCACCATCCACCTCTCCCCTCCCCAGGTCCC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137071 ATCTTTGAAGCGTTTGTTTATCAGCTCTTCCACCATCCACCTCTCCCCTCCCCAGGTCCC 137130

Qy     781 CGATCTAAAATCAAAGAGATTGATTTAGGATGGGTGGGTGCCTTGTCTTCTCTCATTGTT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137131 CGATCTAAAATCAAAGAGATTGATTTAGGATGGGTGGGTGCCTTGTCTTCTCTCATTGTT 137190

Qy     841 CGACATTTTAGTTACGTTTTCTCTGAGCTCTCTGGAAAGCATAAAAGTATAATATCTGTT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137191 CGACATTTTAGTTACGTTTTCTCTGAGCTCTCTGGAAAGCATAAAAGTATAATATCTGTT 137250

Qy     901 AAAAGTTGGATGAATGAACTAATGAACGCAATGGGATTCCAGAAAACTCTGCGGGAGATG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137251 AAAAGTTGGATGAATGAACTAATGAACGCAATGGGATTCCAGAAAACTCTGCGGGAGATG 137310

Qy     961 GGCTAGAGGACGAGGAGGAGGTGGATGAATCAGCCATGTTAGAGAGCCTGGGAAGGTGAG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137311 GGCTAGAGGACGAGGAGGAGGTGGATGAATCAGCCATGTTAGAGAGCCTGGGAAGGTGAG 137370

Qy    1021 CAGAGTTGAAAACTTGATAGATCTAATAATTTACTGGCTCTGGGTTTGTCAGTCACTACA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137371 CAGAGTTGAAAACTTGATAGATCTAATAATTTACTGGCTCTGGGTTTGTCAGTCACTACA 137430

Qy    1081 TTGCAGCAAATGAGATTAGAGCATAGTTGTGGGAGGGAAGGAGGTGACGCAGCAATCTAT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137431 TTGCAGCAAATGAGATTAGAGCATAGTTGTGGGAGGGAAGGAGGTGACGCAGCAATCTAT 137490

Qy    1141 TTGCACCTAGAAATTTTAGGCAAGTGATAGCTGCGTAATCATACTGCGGCACCGTTTTTT 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137491 TTGCACCTAGAAATTTTAGGCAAGTGATAGCTGCGTAATCATACTGCGGCACCGTTTTTT 137550

Qy    1201 TCTTGCAGCAGTAGCTGCTTGCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137551 TCTTGCAGCAGTAGCTGCTTGCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCG 137610

Qy    1261 GCTCTCTCCTGCAGGATCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGTC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137611 GCTCTCTCCTGCAGGATCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGTC 137670

Qy    1321 GCCACTGCTTCAGTCCGGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTG 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137671 GCCACTGCTTCAGTCCGGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTG 137730

Qy    1381 GATCGTCGTTCCCTCAGTCCGGACGCCACTGCTAGGTCCGACCACCGCCGCTTCTGATAT 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137731 GATCGTCGTTCCCTCAGTCCGGACGCCACTGCTAGGTCCGACCACCGCCGCTTCTGATAT 137790

Qy    1441 TTCGGTGAGTCTTTTCCTGTGGAGGTTTGGTCTCCCGATCTCTGTGGTAGCCACCTTAGG 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137791 TTCGGTGAGTCTTTTCCTGTGGAGGTTTGGTCTCCCGATCTCTGTGGTAGCCACCTTAGG 137850

Qy    1501 CGTGTACGGTCCTTTGAAAA 1520
          |||||||||||||||||||||
Db  137851 CGTGTACGGTCCTTTGAAAA 137870
```

Db 1501 CGTGTACGGTCCTTTGAAAA 1520

RESULT 3
AC004074
LOCUS AC004074 153578 bp DNA linear PRI 12-APR-1998
DEFINITION Homo Sapiens Chromosome X clone bWXD759, complete sequence.
ACCESSION AC004074
VERSION AC004074.1 GI:3046270
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 153578)
AUTHORS Chen,E., Brownstein,B.H., States,D.J., Schlessinger,D. and Mazzarella,R.
TITLE Direct Submission
JOURNAL Unpublished (1997)
REFERENCE 2 (bases 1 to 153578)
AUTHORS Brownstein,B.H., States,D.J. and Mazzarella,R.
TITLE Direct Submission
JOURNAL Submitted (29-JAN-1998) Center for Genetics in Medicine, Box 8232, Washington University School of Medicine, 4566 Scott Avenue, St. Louis, MO 63110, USA
REFERENCE 3 (bases 1 to 153578)
AUTHORS Brownstein,B.H., States,D.J. and Mazzarella,R.
TITLE Direct Submission
JOURNAL Submitted (12-APR-1998) Center for Genetics in Medicine, Box 8232, Washington University School of Medicine, 4566 Scott Avenue, St. Louis, MO 63110, USA
COMMENT On Apr 13, 1998 this sequence version replaced gi:2822135.
Current status of this project is available at:
'http://www.ibc.wustl.edu/cgm/seq_projects.html'
Submitted by:
Ellson Chen,
Advanced Center for Genetic Technology,
Applied Biosystems Division of Perlin Elmer Corp.,
850 Lincoln Center Drive,
Foster City, CA 94404 USA
e-mail: ellson@genseq.apldbio.com

and

Buddy Brownstein,
Center for Genetics in Medicine,
Washington University School of Medicine, Box 8232
4566 Scott Avenue,
St. Louis, MO 63110, USA
e-mail: buddy@genetics.wustl.edu

and

David J. States,
Institute for Biomedical Computing
Washington University in St. Louis
700 South Euclid Ave.
St. Louis, MO 63108 USA
e-mail: states@ibc.wustl.edu.

FEATURES Location/Qualifiers
source 1. .153578
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="bWXD759"

BASE COUNT 45069 a 29281 c 29096 g 50132 t
ORIGIN

Query Match 100.0%; Score 1520; DB 9; Length 153578;
Best Local Similarity 100.0%; Pred. No. 5.4e-312;
Matches 1520; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ACTTAAAGGAAAAATTTATCTATAAACTGACAGAATTTAGAAATAAATACAACAATATGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136351 ACTTAAAGGAAAAATTTATCTATAAACTGACAGAATTTAGAAATAAATACAACAATATGT 136410

Qy      61 AAACAGTTTTAATATCTGTGATAGTAACAAATTCTTTAAAATCTGGAAAATAATAGTCACT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136411 AAACAGTTTTAATATCTGTGATAGTAACAAATTCTTTAAAATCTGGAAAATAATAGTCACT 136470

Qy     121 TAAAATTTTAAAAAAATTGTTCAATTAATAAATGATCCAAGTTAGAAATATGAACAAAATA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136471 TAAAATTTTAAAAAAATTGTTCAATTAATAAATGATCCAAGTTAGAAATATGAACAAAATA 136530

Qy     181 AACCTCACCAATAATTACTATAGAGAGGAAAATTTTAATTACTGCAAAGCTTTCCATCCTA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136531 AACCTCACCAATAATTACTATAGAGAGGAAAATTTTAATTACTGCAAAGCTTTCCATCCTA 136590

Qy     241 TAAATACATTATCAAATAGTTTAACCATTTCTTTAATGCTGAGATTTAGATTATTTCCAA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6591 TAAATACATTATCAAATAGTTTAACCATTTCTTTAATGCTGAGATTTAGATTATTTCCAA 136650

Qy     301 TTAACTCAAAAGCATCAAGCAAATGTTATGATTTCTAAGAATAAACATAACTTTCCATTT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136651 TTAACTCAAAAGCATCAAGCAAATGTTATGATTTCTAAGAATAAACATAACTTTCCATTT 136710

Qy     361 TGGCTTTTGTATATATGTATATTTCTAACGGCTGTTAAAGCCAGCATTAAGAAGGAGAAG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136711 TGGCTTTTGTATATATGTATATTTCTAACGGCTGTTAAAGCCAGCATTAAGAAGGAGAAG 136770

Qy     421 CAGAAAGTCAGTATTGGGACTGGGGTTATTTATAAGCCAGGCAACTGGTTAATTGTGGTT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136771 CAGAAAGTCAGTATTGGGACTGGGGTTATTTATAAGCCAGGCAACTGGTTAATTGTGGTT 136830

Qy     481 AATTGTCTGGTATGTTTACTAGTCACGTAGTTGTATACACCATACTAGTTTTTCATCACA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136831 AATTGTCTGGTATGTTTACTAGTCACGTAGTTGTATACACCATACTAGTTTTTCATCACA 136890

Qy     541 GGCCCTCATTCGCCCCCACTGCCATCGGACTTCCTCCTCCTCCCCTCACAGGAAATGTTT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136891 GGCCCTCATTCGCCCCCACTGCCATCGGACTTCCTCCTCCTCCCCTCACAGGAAATGTTT 136950

Qy     601 CGAGAATTTTTCAACCTAAAATCATATAGCTTGTGAAAAATACCGACAAACATAATATAG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  136951 CGAGAATTTTTCAACCTAAAATCATATAGCTTGTGAAAAATACCGACAAACATAATATAG 137010

Qy     661 AATATTTAAATAACTGACACGCCACCTAAAGACCATCAGTGCTAATTCCTGGTGTTTTTA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137011 AATATTTAAATAACTGACACGCCACCTAAAGACCATCAGTGCTAATTCCTGGTGTTTTTA 137070

Qy     721 ATCTTTGAAGCGTTTGTTTATCAGCTCTTCCACCATCCACCTCTCCCCTCCCCAGGTCCC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137071 ATCTTTGAAGCGTTTGTTTATCAGCTCTTCCACCATCCACCTCTCCCCTCCCCAGGTCCC 137130

Qy     781 CGATCTAAAATCAAAGAGATTGATTTAGGATGGGTGGGTGCCTTGTCTTCTCTCATTGTT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137131 CGATCTAAAATCAAAGAGATTGATTTAGGATGGGTGGGTGCCTTGTCTTCTCTCATTGTT 137190

Qy     841 CGACATTTTAGTTACGTTTTCTCTGAGCTCTCTGGAAAGCATAAAAGTATAATATCTGTT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137191 CGACATTTTAGTTACGTTTTCTCTGAGCTCTCTGGAAAGCATAAAAGTATAATATCTGTT 137250

Qy     901 AAAAGTTGGATGAATGAACTAATGAACGCAATGGGATTCCAGAAAACTCTGCGGGAGATG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137251 AAAAGTTGGATGAATGAACTAATGAACGCAATGGGATTCCAGAAAACTCTGCGGGAGATG 137310

Qy     961 GGCTAGAGGACGAGGAGGAGGTGGATGAATCAGCCATGTTAGAGAGCCTGGGAAGGTGAG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137311 GGCTAGAGGACGAGGAGGAGGTGGATGAATCAGCCATGTTAGAGAGCCTGGGAAGGTGAG 137370

Qy    1021 CAGAGTTGAAAACTTGATAGATCTAATAATTTACTGGCTCTGGGTTTGTCAGTCACTACA 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137371 CAGAGTTGAAAACTTGATAGATCTAATAATTTACTGGCTCTGGGTTTGTCAGTCACTACA 137430

Qy    1081 TTGCAGCAAATGAGATTAGAGCATAGTTGTGGGAGGGAAGGAGGTGACGCAGCAATCTAT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137431 TTGCAGCAAATGAGATTAGAGCATAGTTGTGGGAGGGAAGGAGGTGACGCAGCAATCTAT 137490

Qy    1141 TTGCACCTAGAAATTTTAGGCAAGTGATAGCTGCGTAATCATACTGCGGCACCGTTTTTT 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137491 TTGCACCTAGAAATTTTAGGCAAGTGATAGCTGCGTAATCATACTGCGGCACCGTTTTTT 137550

Qy    1201 TCTTGCAGCAGTAGCTGCTTGCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137551 TCTTGCAGCAGTAGCTGCTTGCGGAGGAGGTCTGCCCACTGCAGCTCTCTGCAGTCTCCG 137610

Qy    1261 GCTCTCTCCTGCAGGATCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGTC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137611 GCTCTCTCCTGCAGGATCGGTCAACGCAGCCGTCGCCGCCCTCTGCACCCAGCCCAGGTC 137670

Qy    1321 GCCACTGCTTCAGTCCGGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTG 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137671 GCCACTGCTTCAGTCCGGTTCTCAAAGCCTCAGCACCATCTTTTATCCCCGAGCAGCCTG 137730

Qy    1381 GATCGTCGTTCCCTCAGTCCGGACGCCACTGCTAGGTCCGACCACCGCCGCTTCTGATAT 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137731 GATCGTCGTTCCCTCAGTCCGGACGCCACTGCTAGGTCCGACCACCGCCGCTTCTGATAT 137790

Qy    1441 TTCGGTGAGTCTTTTCCTGTGGAGGTTTGGTCTCCCGATCTCTGTGGTAGCCACCTTAGG 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  137791 TTCGGTGAGTCTTTTCCTGTGGAGGTTTGGTCTCCCGATCTCTGTGGTAGCCACCTTAGG 137850

Qy    1501 CGTGTACGGTCCTTTGAAAA 1520
          ||||||||||||||||||||
Db  137851 CGTGTACGGTCCTTTGAAAA 137870
```

RESULT 4
AX327801